2

PHYLOGENETIC SYSTEMATICS

OVERVIEW AND GOALS

As introduced in the previous chapter, **phylogeny** refers to the evolutionary history or pattern of descent of a group of organisms and is one of the primary goals of systematics. **Phylogenetic systematics**, or **cladistics**, is that branch of systematics concerned with inferring phylogeny. Ever since Darwin laid down the fundamental principles of evolutionary theory, one of the major goals of the biological sciences has been the determination of life's history of descent. This phylogeny of organisms, visualized as a branching pattern, can be determined by an analysis of characters from living or fossil organisms, utilizing phylogenetic principles and methodology.

As reviewed in Chapter 1, a phylogeny is commonly represented in the form of a **cladogram**, or **phylogenetic tree**, a branching diagram that conceptually represents an estimate of phylogeny (Figure 2.1). The lines of a cladogram are known as **lineages**, often referred to simply as "branches." Lineages represent the sequence of ancestral-descendant populations through time, ultimately denoting descent. (The term "lineage" is treated here as a single branch; "clade" is defined as a given common ancestor plus all descendants, including two or more lineages, being essentially equivalent to a monophyletic group; see later discussion.)

Evolution may occur within lineages over time and is recognized as a change from a preexisting **ancestral** (also called **plesiomorphic** or **primitive**) condition to a new, **derived**

10.1016/B978-0-12-374380-0.50002-6

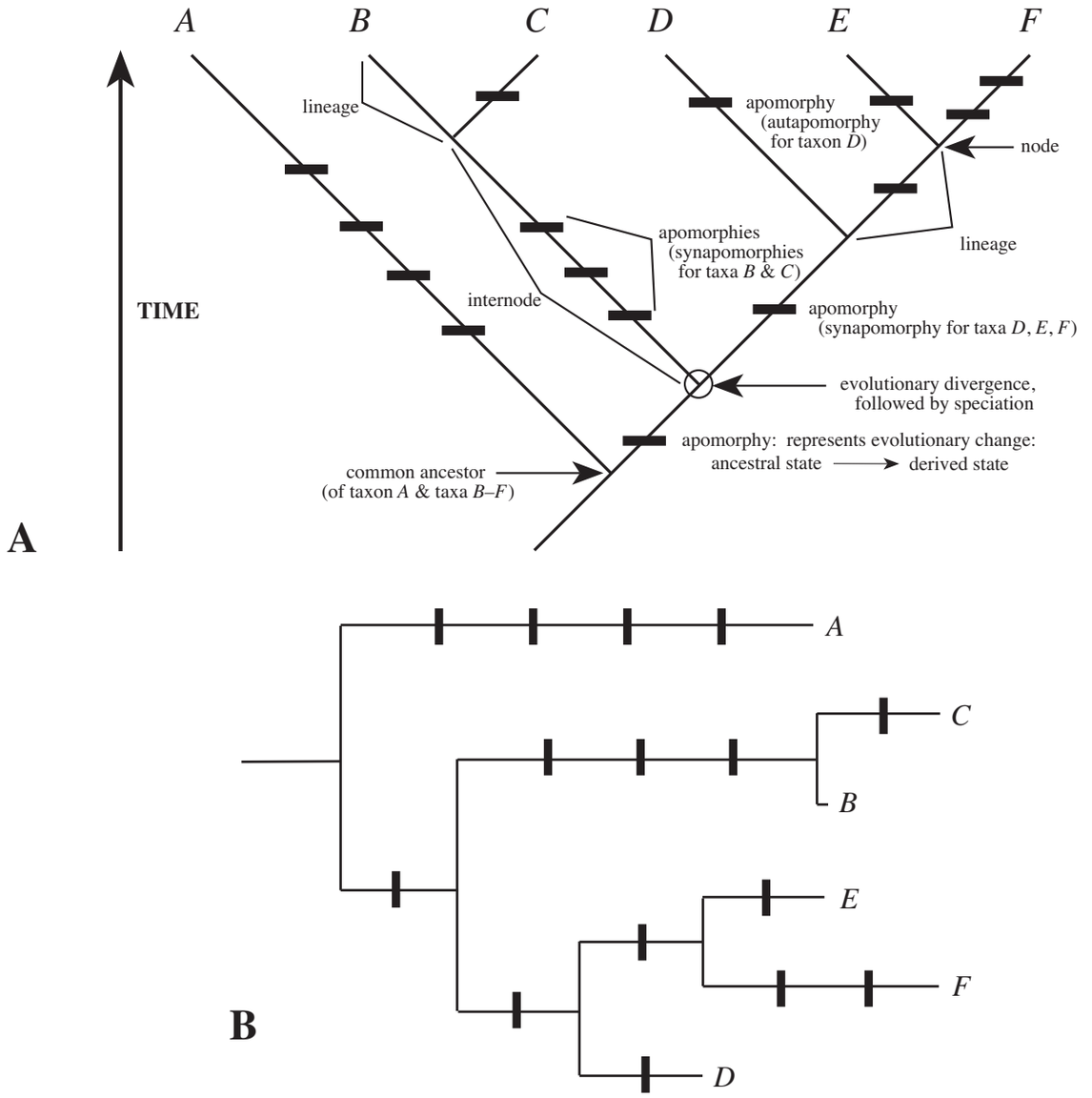


FIGURE 2.1 **A.** Example of a cladogram or phylogenetic tree for taxa *A–F*, with apomorphies indicated by thick hash marks. See text for explanation of terms. **B.** Same cladogram topology but drawn horizontally, with branch lengths scaled to number of apomorphic changes.

(also called **apomorphic** or **advanced**) condition. The derived condition, or **apomorphy**, represents an evolutionary novelty. As seen in Figure 2.1A, an apomorphy that unites two or more lineages is known as a **synapomorphy** (*syn*, together); one that occurs within a single lineage is called an **autapomorphy** (*aut*, self). However, either may be referred to simply as an **apomorphy**, a convention used throughout this book.

Any branching of the cladogram represents lineage **divergence** or **diversification**, the formation of two separate lineages from one **common ancestor**. (The two lineages could diverge into what would be designated separate species, the process of forming two species from one termed **speciation**.) The *point* of divergence of one clade into two (where the most recent common ancestor of the two divergent clades is located) is termed a **node**; the region between two nodes is called an **internode** (Figure 2.1A). Cladograms may be represented in different ways. Figure 2.1B shows the same cladogram as in Figure 2.1, but shifted 90° clockwise and with the lineages drawn perpendicular to one another and of

a length reflective of the number of apomorphic changes. (These two representations of a cladogram have the same **topology**, which is the structure of the branching diagram, i.e., how lineages, including those terminating in taxa, are connected together.)

Cladograms have an implied, but *relative*, time scale. For example, in Figure 2.1, the common ancestor giving rise to taxa E and F occurred later in time than that giving rise to taxa D, E, and F, but we do not know when the lineage splitting at these nodes occurred or how long the lineages are in terms of real time. The term **phylogram** is often used for a cladogram that has an *absolute* time scale, such that nodes and branch lengths are calibrated and correspond more closely to real elapsed time. (See later discussion.)

Why study phylogeny? Knowing the pattern of descent, in the form of a cladogram, can be viewed as an important end in itself. The branching pattern derived from a phylogenetic analysis may be used to infer the collective evolutionary changes that have occurred in ancestral/descendant populations through time. Thus, a knowledge of phylogenetic relationships may be invaluable in understanding structural evolution as well as in gaining insight into the possible functional, adaptive significance of hypothesized evolutionary changes. The cladogram can also be used to classify life in a way that directly reflects evolutionary history. Cladistic analysis may also serve as a tool for inferring biogeographic and ecological history, assessing evolutionary processes, and making decisions in the conservation of threatened or endangered species. (See Chapter 19.)

The principles, methodology, and applications of phylogenetic analyses are described in the remainder of this chapter.

TAXON SELECTION

The study of phylogeny begins with the selection of **taxa** (taxonomic groups) to be analyzed, which may include living and/or fossil organisms. Taxon selection includes both the group as a whole, called the study group or **ingroup**, and the individual unit taxa, often termed **O**perational **T**axonomic **U**nits, or **OTUs**. The rationale as to *which* taxa are selected from among many rests by necessity on previous classifications or phylogenetic hypotheses. The ingroup is often a traditionally defined taxon for which there are competing or uncertain classification schemes, the objective being to test the bases of those different classification systems or to provide a new classification system derived from the phylogenetic analysis. The OTUs are previously classified members of the study group and may be species or taxa consisting of groups of species (e.g., traditional genera or families). Sometimes named subspecies or even populations, if distinctive and presumed to be on their own evolutionary track, can be used as OTUs in a cladistic analysis.

In addition, outgroup OTUs are selected. An **outgroup** is a taxon that is closely related to but not a member of the ingroup (see **Outgroup Comparison**). Outgroups are used to "root" a tree (see later discussion).

Some caution should be taken in choosing which taxa to study. First, the OTUs must be well-circumscribed and delimited from one another. Second, the study group itself should be large enough so that all probable closely related OTUs are included in the analysis. Stated strictly, both OTUs and the ingroup as a whole must be assessed for **monophyly** before the analysis is begun (see below). In summary, the initial selection of taxa in a cladistic analysis, both study group and OTUs, should be questioned beforehand to avoid the bias of blindly following past classification systems.

CHARACTER ANALYSIS

DESCRIPTION

Fundamental in any systematic study is description, the characterization of the attributes or features of taxa using any number of types of evidence (see Chapters 9–14). A systematist may make original descriptions (for example, acquisition of DNA sequence data) or rely partly or entirely on previously published or cataloged data. In any case, it cannot be overemphasized that the ultimate validity of a phylogenetic study depends on the descriptive accuracy and completeness of the primary investigator. Thorough research and a comprehensive familiarity with the literature on the taxa and characters of concern are prerequisites to a phylogenetic study.

CHARACTER SELECTION AND DEFINITION

After taxa are selected and the basic research and literature survey are completed, the next step in a phylogenetic study is the actual selection and definition of **characters** and **character states** from the descriptive data. (Recall that a character is an attribute or feature; character states are two or more forms of a character.) Generally, those features that (1) are genetically determined and heritable (termed "intrinsic"), (2) are relatively invariable within an OTU, and (3) denote clear discontinuities from other similar characters and character states should be utilized. However, the selection of a finite number of characters from the virtually infinite number that could be used adds an element of subjectivity to the study. Thus, it is important to realize that any analysis is inherently biased simply by *which*

characters are selected and *how* the characters and character states are defined. (In some cases, certain characters may be weighted over others; see later discussion.)

Characters used in phylogenetic analyses are usually conceptually divided into two classes: "**morphological**," essentially equivalent to nonmolecular features, such as organ morphology (Chapter 9), anatomy (Chapter 10), embryology (Chapter 11), palynology (Chapter 12), and some aspects of reproductive biology (Chapter 13); and "**molecular**," derived from genetic data, such as DNA sequences (see Chapter 14).

Morphological features are generally the manifestation of numerous intercoordinated genes, and because evolution occurs by a change in one or more of those genes, the precise definition of a morphological feature in terms of characters and character states may be problematic. A structure may be defined broadly as a whole entity with several components. Alternatively, discrete features of a structure may be defined individually as separate characters and character states. For example, in comparing the evolution of fruit morphology within some study group, the character "fruit type" might be designated as two character states: berry versus capsule, *or* the characteristics of the fruit may be subdivided into a host of characters with their corresponding states, for example, "fruit shape," "fruit wall texture," "fruit dehiscence," and "seed number." (These characters may be correlated, however; see later discussion.) In practice, characters are divided only enough to communicate differences between two or more taxa. However, this type of terminological atomization may be misleading with reference to the effect of specific genetic changes in evolution, as genes do not normally correspond one for one with taxonomic characters. The morphology of a structure is the end product of development, involving a host of complex interactions of the entire genotype.

Molecular characters may be less "subjective" than morphological ones, but they are not fool proof. Polymorphisms or uncertainties in base determination may occur for DNA sequence data. Sequence alignment, in particular, may not be clear-cut if sequences between taxa are very different (e.g., some taxa having significant deletions or insertions), necessitating often "black-box" sequence alignment programs. And the possibility of paralogy due to ancestral gene duplication or hybridization may confound comparison of sequences that are homologous. (See Chapter 14.)

CHARACTER STATE DISCRETENESS

Because phylogenetic systematics entails the recognition of an evolutionary transformation from one state to another, an important requirement of character analysis is that character states be discrete or discontinuous from one another. Molecular characters and their states are usually discrete (see Chapter 14), although polymorphism of nucleotide base sites can occur. For some morphological, qualitative characters such as corolla color, the discontinuity of states is clear; e.g., the corolla is yellow in some taxa and blue in others. But for other features, character states may not actually be clearly distinguishable from one another. This lack of discontinuity often limits the number of available characters and is often the result of variation of a feature either within a taxon or between taxa. Because character states must be clearly discrete from one another in order to be used in a cladistic analysis, they must be evaluated for discontinuity. A standard way to evaluate state discontinuity is to do a statistical analysis, e.g., by comparing the means, ranges, and standard deviations of each character for all taxa in the analysis (including outgroup taxa; see later discussion). Such an analysis may reveal two or more classes of features that may be defined as discrete character states (Figure 2.2). The investigator must decide what constitutes discreteness, such as lack of overlap of ranges or lack of overlap of ±1 standard deviation. Additional statistical tests, such as ANOVAS, *t*-tests, or multivariate statistics, may be used as other criteria for evaluating character state discontinuity. (See Appendix 4 for details.)

CHARACTER CORRELATION

Another point to consider in character selection and definition (generally with nonmolecular data) is whether there is possible correlation of characters. Character correlation is an interaction between what are defined as separate characters, but which are actually components of a common structure, the manifestation of a single evolutionary novelty. Two or more characters are correlated if a change in one always accompanies a corresponding change in the other, bringing up the possibility that they are genetically linked. When characters defined in a cladistic analysis are correlated, including them in the analysis (as two or more separate characters) may inadvertently weight what could otherwise be listed as a single character. In the example above, in which the original single character "fruit type" is subdivided into many characters ("fruit shape," "fruit wall texture," "fruit dehiscence," and "seed number"), it is likely that these separate characters are correlated with an evolutionary shift from one fruit type (e.g., "capsule") to another (e.g., "berry"). This is tested simply by determining if there is any variation in the character states of the subdivided characters between taxa. If characters appear to be correlated, they should either be combined into a single character or *scaled*, such that each component character gets a reduced weight in a phylogenetic analysis (see **Character Weighting**, page 23).

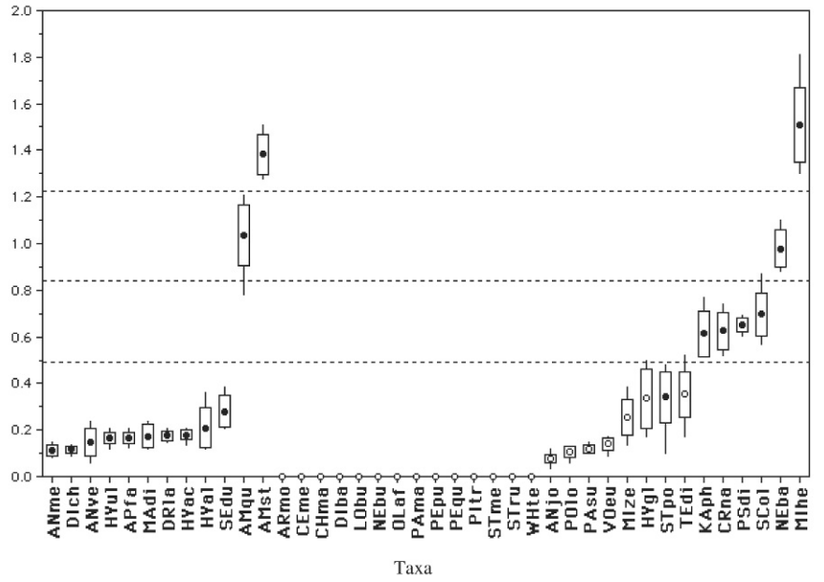


FIGURE 2.2 Example of a pollen character (exine wall foot-layer thickness) for which the character states are quantitatively analyzed for each taxon. The dashed horizontal lines represent "breaks" or discontinuities between states. Solid dots are means, vertical lines are ranges, and boxes are ±1 standard deviation from the mean (used here as the measure of "discreteness"). Outgroup taxa are to the left, ingroup taxa to the right. (From Levin, G. A., and M. G. Simpson. 1994. Annals of the Missouri Botanical Garden 81: 203–238.)

HOMOLOGY ASSESSMENT

One concept critical to cladistics is that of **homology**, which can be defined as similarity resulting from common ancestry. Characters or character states of two or more taxa are homologous if those same features were present in the common ancestor of the taxa. For example, the flower of a daisy and the flower of an orchid are homologous as flowers because their common ancestor had flowers, which the two taxa share by continuity of descent. Taxa with homologous features are presumed to share, by common ancestry, the same or similar DNA sequences or gene assemblages that may, e.g., determine the development of a common structure such as a flower. (Unfortunately, molecular biologists often use the term *homology* to denote similarity in DNA sequence, even though the common ancestry of these sequences may not have been tested; using the term *sequence similarity* in this case is preferred.)

Homology may also be defined with reference to similar structures within the same individual; two or more structures are homologous if the DNA sequences that determine their similarity share a common evolutionary history. For example, carpels of flowering plants are considered to be homologous with leaves because of a basic similarity between the two in form, anatomy, and development. Their similarity may be hypothesized to be the result of a "sharing" of common genes (or of duplicated genes) that direct their development. The duplication and subsequent divergence of genes is a type of intra-individual or intra-species homology; the genes are similar because of origin from a common ancestor, in this case the gene prior to duplication.

Molecular data must also be assessed for homology. Generally, a common nucleotide base site for a given gene or intergenic region is assumed to be homologous among the OTUs of a study. However, gene duplication or past hybridization (e.g., resulting in polyploidy; see Chapter 13) may confound homology of even DNA sequence data, in that nonhomologous, paralogous genes or sequences are unknowingly being compared (Chapter 14).

Similarity between taxa can arise not only by common ancestry, but also by independent evolutionary origin. Similarity that is *not* the result of homology is termed **homoplasy** (also sometimes termed *analogy*). Homoplasy may arise in two ways: convergence (equivalent to "parallelism," here) or reversal. **Convergence** is the independent evolution of a similar feature in two or more lineages. Thus, liverwort gametophytic leaves and lycophyte sporophytic leaves evolved independently as dorsiventral, photosynthetic appendages; their similarity is homoplasious by convergent evolution. (However, although "leaves" in the two groups evolved independently, they could possibly be homologous in the sense of utilizing at least some gene complexes of common origin that function in the development of bifacial organs. This is unknown at present.)

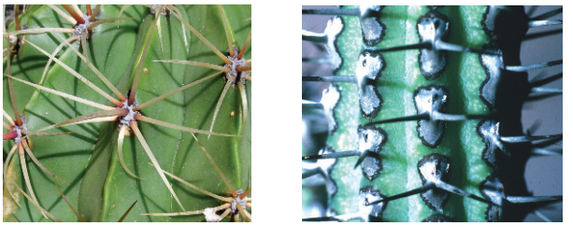


FIGURE 2.3 Comparison of cactus (left) and euphorb (right) spines, which are not homologous as spines.

Reversal is the loss of a derived feature with the reestablishment of an ancestral feature. For example, the reduced flowers of many angiosperm taxa, such as *Lemna*, lack a perianth; comparative and phylogenetic studies have shown that flowers of these taxa lack the perianth by secondary loss, i.e., via a reversal, reverting to a condition prior to the evolution of a reproductive shoot having a perianth-like structure.

The determination of homology is one of the most challenging aspects of a phylogenetic study and may involve a variety of criteria. Generally, homology is hypothesized based on some evidence of similarity, either direct similarity (e.g., of structure, position, or development) or similarity via a gradation series (e.g., intermediate forms between character states). Homology should be assessed for each character of all taxa in a study, particularly of those taxa having *similarly termed* character states. For example, both the cacti and stem-succulent euphorbs have spines (Figure 2.3). Thus, for the character "spine presence/absence," the character state "spines present" may be assigned to both of these two taxa in a broad cladistic analysis. Whether intended or not, this designation of the same character state for two or more taxa presupposes that these features are homologous in those taxa and arose by common evolutionary origin. Thus, a careful distinction should be made between terminological similarity and similarity by homology. In the above example, more detailed study demonstrates that the spines of cacti and euphorbs are quite different in origin, cacti having leaf spines arising from an areole (a type of short shoot), euphorbs having spines derived from modified stipules. Despite the similarity between spines of cacti and stem-succulent euphorbs, their structural and developmental *dissimilarity* indicates that they are homoplasious and had independent evolutionary origins (with similar selective pressures, i.e., protection from herbivores). This hypothesis necessitates a redefinition of the characters and character states, such that the two taxa are not coded the same.

Hypotheses of homology are tested by means of the cladistic analysis. The *totality* of characters are used to infer the most likely evolutionary tree, and the original assessment of homology is checked by determining if convergences or reversals must be invoked to explain the distribution of character states on the final cladogram (see later discussion).

CHARACTER STATE TRANSFORMATION SERIES AND POLARITY

After the characters and character states have been selected and defined and their homologies have been assessed, the character states for each character are arranged in a sequence, known as a **transformation series** or **morphocline**. Transformation series represent the hypothesized sequence of evolutionary change, from one character state to another, in terms of direction and probability. For a character with only two character states, known as a **binary character**, obviously only one transformation series exists. For example, for the character "ovary position" having the states "inferior" and "superior," the implied transformation series is "inferior ⇔ superior." This two-state transformation series represents (at least initially) a single, hypothesized evolutionary step, the direction of which is unspecified, being either "inferior ⇒ superior" or "superior ⇒ inferior."

Characters having three or more character states, known as **multistate characters**, can be arranged in transformation series that are either ordered or unordered. An **unordered** transformation series allows for each character state to evolve into every other character state with equal probability, i.e., in a single evolutionary step. For example, an unordered transformation series for a three-state character is shown in Figure 2.4A; one for a four-state character is shown in Figure 2.4B and C. An **ordered** transformation series places the character states in a predetermined sequence that may be linear (Figure 2.4D) or branched (Figure 2.4E). Ordering a transformation series limits the direction of character state changes. For example, in Figure 2.4E, the evolution of "2 stamens" from "5 stamens" (or vice versa) takes two evolutionary steps and necessitates passing through the intermediate condition, "4 stamens"; the comparable unordered series takes a single step between "2 stamens" and "5 stamens" (and between all other character states; Figure 2.4B).

The rationale for an ordered series is the assumption or hypothesis that evolutionary change proceeds gradually, such that going from one extreme to another most likely entails passing through some recognizable intermediate condition. Ordered transformation series are generally postulated vis-à-vis some obvious intergradation of character states or stages in the ontogeny of a character. A general suggestion in cladistic analyses is to code all characters as unordered unless there is compelling evidence for an ordered transformation, such as the presence of a vestigial feature in a derived structure.

A final aspect of character state transformations is the assignment of polarity. **Polarity** is the designation of relative

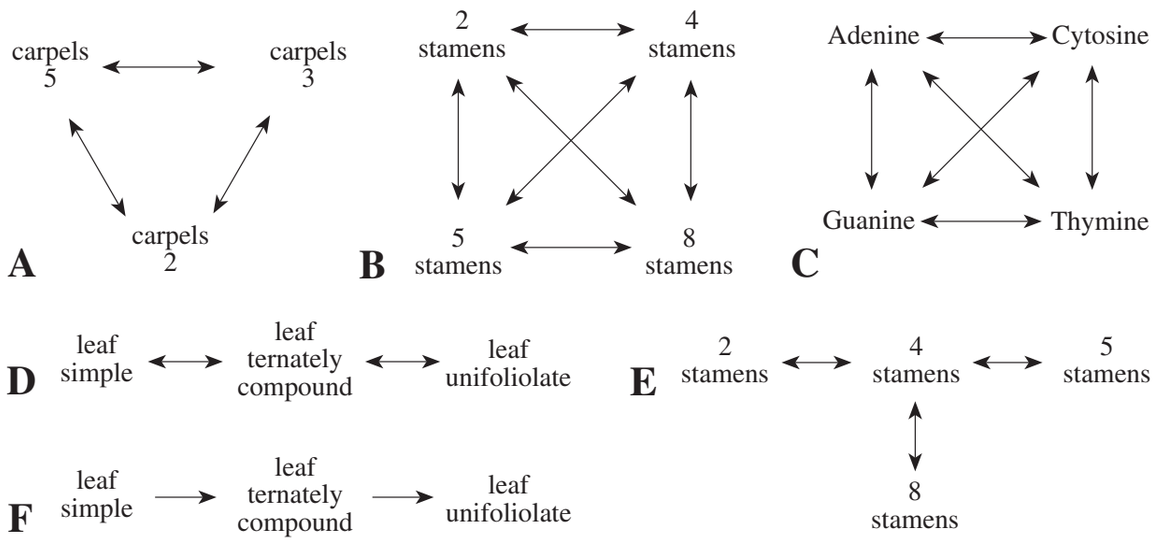


FIGURE 2.4 Examples of character state transformations used in a cladistic analysis. **A.** Unordered, three-state character. **B,C.** Unordered, four-state character. **D.** Ordered, three-state character. **E.** Ordered, four-state character. **F.** Ordered and polarized, three-state character.

ancestry to the character states of a morphocline. As summarized earlier, a change in character state represents a heritable evolutionary modification from a preexisting structure or feature (termed **plesiomorphic**, **ancestral**, or primitive) to a new structure or feature (**apomorphic**, **derived**, or advanced). For example, for the character "ovary position," with character states "superior" and "inferior," if a superior ovary is hypothesized as ancestral, the resultant "polarized" morphocline would be "superior ⇒ inferior." For a multistate character (e.g., "leaf type" in Figure 2.4D), an example of a polarized, ordered transformation series is seen in Figure 2.4F. (In this example, the unifoliolate leaf possesses a vestige of an ancestrally compound condition, evidence that it should terminate in the ordered character state transformation.) In practice polarity determination of characters is usually attained by assigning one or more outgroups (see **Outgroup Comparison**, page 33).

CHARACTER WEIGHTING

As part of a phylogenetic analysis, the investigator may choose to weight characters. Character **weighting** is the assignment of greater or lesser taxonomic importance to certain characters over other characters in determining phylogenetic relationships. Assigning a character greater weight has the effect of listing it more than once in the character × taxon matrix (see later section) in order to possibly "override" competing changes in unweighted characters. (Note that fractional weights can also be assigned using computer algorithms.)

Characters may be given greater weight in cases for which the designation of homology is considered relatively certain. The expectation is that, by increasing the weight of characters for which homoplasy is deemed unlikely, taxa will be grouped by real, shared derived features. Such characters given greater weight may be hypothesized as having homologous states for various reasons. For example, a feature distinctive for two or more taxa may be structurally or developmentally complex, such that the independent evolution of the same character state would seem very unlikely. (It should be realized, however, that if a feature is most likely highly adaptive, convergence of similar complex features in two or more taxa may not necessarily be ruled out.)

Characters may be weighted unintentionally because they are correlated, i.e., the corresponding character state values of two or more characters are always present in all taxa and believed to be aspects of the same evolutionary novelty. In order to prevent excess weighting of correlated characters, they may be **scaled**, meaning that each character receives a weight that is the inverse of the number of characters (e.g., if there are three correlated characters, each receives a weight of 1/3).

In practice, character weighting of morphological data is rarely done, in part because of the arbitrariness of determining the amount of weight a character or state should have. A frequent exception, however, is molecular data, for which empirical studies may justify the rationale for and degree of weighting. *Evolutionary models* utilize a sophisticated type of character weighting (see later discussion).

Alternatively, weighting may be done *after* the first stage of a phylogenetic analysis. Those characters that exhibit reversals or parallelisms on the cladogram are recognized and

A

	0	1	2	3
0	0	1	2	3
1	1	0	1	2
2	2	1	0	1
3	3	2	1	0

B

	0	1	2	3
0	0	1	1	1
1	1	0	1	1
2	1	1	0	1
3	1	1	1	0

C

	0	1	2	3
0	0	1	2	3
1	∞	0	1	2
2	∞	∞	0	1
3	∞	∞	∞	0

D

	0	1	2	3
0	0	1	5	5
1	1	0	5	5
2	5	5	0	1
3	5	5	1	0

FIGURE 2.5 Character step matrices for: **A.** Ordered character. **B.** Unordered character. **C.** Irreversible character. **D.** Differentially weighted character.

given less weight over those that do not, sometimes as a direct function of the degree of homoplasy they exhibit. For example, if, after a cladistic analysis, a character exhibits two convergent changes, that character would be given a weight of 1/2 in a second cladistic analysis. This type of *a posteriori* analysis is called *successive weighting* (which relies on the assumption that the initial tree(s) are close to an accurate representation of phylogeny). Often, the rescaled consistency index (RC) value is used as a basis for successive weighting (see **Measures of Homoplasy**, page 39).

CHARACTER STEP MATRIX

As reviewed earlier, assigning a character state transformation determines the number of steps that may occur when going from one character state to another. Computerized phylogeny reconstruction algorithms available today permit a more precise tabulation of the number of steps occurring between each pair of character states through a **character step matrix**. The matrix consists of a listing of character states in a top row and left column; intersecting numbers within the matrix indicate the number of steps required, going from states in the left column to states in the top row. For example, the character step matrix of Figure 2.5A illustrates an ordered character state transformation series, such that a single step is required when going from state 0 to state 1 (or state 1 to state 0), two steps are required when going from state 0 to state 2, etc. The character step matrix of Figure 2.5B shows an unordered transformation series, in which a single step is required when going from one state to any other (nonidentical) state. Character step matrices need not be symmetrical; that of Figure 2.5C illustrates an ordered transformation series but one that is irreversible, disallowing a change from a higher state number to a lower state number (e.g., from state 2 to state 1) by requiring a large number of step changes (symbolized by "∞"). Character step matrices are most useful with specialized types of data. For example, the matrix of Figure 2.5D could represent DNA sequence data, where 0 and 1 are the states for the two purines (adenine and guanine) and 2 and 3 are the states for the two pyrimidines (cytosine and thymine; see Chapter 14). Note that in this matrix the change from one purine to another purine or one pyrimidine to another pyrimidine (each of these known as a "transition") requires only one step, being biochemically more probable to occur, whereas a change from a purine to a pyrimidine or from a pyrimidine to a purine (termed a "transversion") is given five steps, being more biochemically less likely. Thus, in a cladistic analysis, the latter change will be given substantially more weight.

DNA sequence data may be transformed in a more complicated *evolutionary model*, based on a number of parameters, such as branch length, codon position, base frequency, or transition/transversion ratio. Such models of evolution are an integral component of maximum likelihood and Bayesian analyses (see later discussion).

CHARACTER × TAXON MATRIX

Prior to cladogram construction, characters and character states for each taxon are tabulated in a **character × taxon matrix**, as illustrated in Figure 2.6A. In order to analyze the data using computer algorithms, the characters and character states must be assigned a numerical value. In doing so, character states are assigned nonnegative integer values, typically beginning with 0. Figure 2.6B shows the numerical coding of the matrix of Figure 2.6A. The states are numerically coded in sequence to correspond with the hypothesized transformation series for that character. For example, for the ordered transformation series "leaf type" of Figure 2.4D,F, the character states "simple," "ternately compound," and "unifoliolate" could be enumerated as 0, 1, and 2. In the character × taxon matrix, polarity is established by including one or more outgroup taxa as part of the character × taxon matrix (as in Figure 2.6A,B) and by subsequently "rooting" the tree by placing the outgroups at the extreme base of the final cladogram (see later discussion).

CLADOGRAM CONSTRUCTION

APOMORPHY

A primary tenet of phylogenetic systematics is that derived character states, or **apomorphies**, that are *shared* between two or more taxa (OTUs) constitute evidence that these taxa possess them because of common ancestry. These shared derived character states, or **synapomorphies**, represent the products of unique evolutionary events that may be used

A

	1 Leaf shape	2 Plant habit	3 Petal number	4 Flower color	5 Stamen number	6 Pollen surface
X. alba	elliptic	shrub	five	red	four	spiny
X. lutea	elliptic	herb	five	red	four	smooth
X. nigra	linear	shrub	four	yellow	two	smooth
X. purpurea	linear	shrub	four	yellow	two	spiny
X. rubens	linear	shrub	four	yellow	four	smooth
OUTGROUP	elliptic	shrub	five	yellow	five	smooth

B

	1	2	3	4	5	6
X. alba	0	0	0	1	1	1
X. lutea	0	1	0	1	1	0
X. nigra	1	0	1	0	2	0
X. purpurea	1	0	1	0	2	1
X. rubens	1	0	1	0	1	0
OUTGROUP	0	0	0	0	0	0

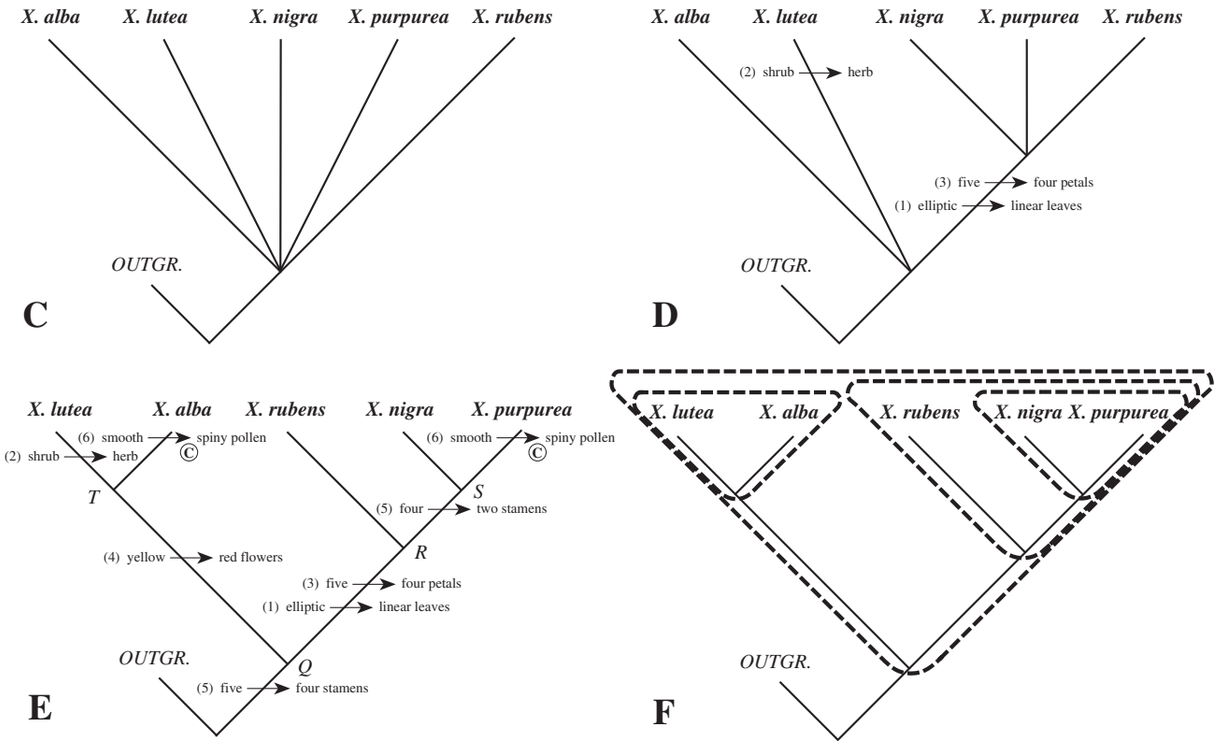


FIGURE 2.6 Character × taxon matrix for five species of hypothetical genus *Xid* plus an outgroup taxon (left column), showing six characters (top row) and their character states (inner columns; character 5 is ordered). **A.** Character state names listed. **B.** Characters and character states converted to numerical values. **C.** Unresolved cladogram. **D.** Addition of characters 1–3. **E.** Most parsimonious cladogram, with addition of other characters. Note common ancestors *Q*, *R*, *S*, *T*, shown for illustrative purposes. Circled **C** denotes convergent homoplasies. **F.** Cladogram at E, with all monophyletic groups circled.

to link two or more taxa in a common evolutionary history. Thus, by sequentially linking taxa together based on their common possession of shared apomorphies, the evolutionary history of the study group can be inferred.

The character × taxon matrix supplies the data for constructing a phylogenetic tree or cladogram. For example, Figure 2.6 illustrates construction of the cladogram for the five species of the hypothetical genus *Xid* from the character × taxon matrix at Figure 2.6A,B. First, the OTUs are grouped together as lineages arising from a single common ancestor above the point of attachment of the outgroup (Figure 2.6C). This unresolved complex of lineages is known as a **polytomy** (see later discussion). Next, *derived* character states are identified and used to sequentially link sets of taxa (Figure 2.6D,E). In this example, synapomorphies include (1) the derived states of characters 1 and 3 that group together *X. nigra*, *X. purpurea*, and *X. rubens*; (2) the derived state of character 4 that groups together *X. alba* and *X. lutea*; (3) the derived state "four stamens" of (ordered) character 5, which is found in all ingroup OTUs and constitutes an apomorphy for the entire ingroup and the derived state "two stamens" of character 5 that groups *X. nigra* and *X. purpurea*. The derived state of character 2 is restricted to the taxon *X. lutea* and is therefore an **autapomorphy**. Autapomorphies occur for a single OTU and are not informative in cladogram construction. Finally, the derived state of character 6 evolved twice, in the lineages leading to both species

X. alba and *X. purpurea*; these independent evolutionary changes constitute homoplasies due to convergence.

One important principle is illustrated in Figure 2.6E for character 5, in which the derived state "four stamens" is an apomorphy for *all* species of the study group, including *X. nigra* and *X. purpurea*. Although the last two species lack the state "four stamens" for that character, they still *share the evolutionary event* in common with the other three species. The lineage terminating in *X. nigra* and *X. purpurea* has simply undergone additional evolutionary change in this character, transforming from four to two stamens (Figure 2.6E).

RECENCY OF COMMON ANCESTRY

Cladistic analysis allows for a precise definition of biological relationship. *Relationship* in phylogenetic systematics is a measure of **recency of common ancestry**. Two taxa are more closely related to one another if they share a common ancestor that is more recent in time than the common ancestor they share with other taxa. For example, in Figure 2.7A taxon *C* is more closely related to taxon *D* than it is to taxon *E* or *F*. This is true because the common ancestor of taxa *C* and *D* is more recent in time (closer to the present) than is the common ancestor of taxa *C, D, E,* and *F* (Figure 2.7A). In the earlier example of Figure 2.6E, it is evident that *X. nigra* and *X. purpurea* are more closely related to one another than either is to *X. rubens*. This is because the first two species together share a common ancestor (*S*) that is more recent in time than the common ancestor (*R*) that they share with *X. rubens*. Similarly *X. rubens* is more closely related to *X. nigra* and *X. purpurea* than it is to either *X. lutea* or *X. alba* because the first three taxa share a common ancestor (*R*) that is more recent in time than *Q*, which is the common ancestor shared by all five species.

The fact that descent is assessed by means of recency of common ancestry gives the rationale that the branches of a given cladogram may be visually rotated around their junction point or "node" (at the common ancestor) with no change in phylogenetic relationships. For example, the cladograms portrayed in Figure 2.8A, 2.8B, and 2.8C are all the same as that in Figure 2.6E, differing only in that the lineages have been rotated about their common ancestors. As discussed earlier, the **topology** of all these cladograms is exactly the same; only the visual structure of branches varies. (Again note that cladograms can be portrayed in different manners, with taxa at the top, bottom, or sides and with lineages drawn as vertical, horizontal, or angled lines, as in Figure 2.8A–C.)

MONOPHYLY

A very important concept in phylogenetic systematics is that of monophyly, or monophyletic groups. As introduced earlier, a **monophyletic group** or **clade** is a group that consists of a common ancestor plus *all* descendants of that ancestor. The rationale for monophyly is based on the concept of recency of common ancestry. Members of a monophyletic group share one or more unique evolutionary events; otherwise, the group could not generally be identified as monophyletic. For example, four monophyletic groups can be delimited from the cladogram of Figure 2.6E; these are circled in Figure 2.6F. In another example, the monophyletic groups of the cladogram of Figure 2.7A are shown in Figure 2.7B. Note that all monophyletic groups include the common ancestor plus all lineages derived from the common ancestor, with each most recent lineage terminating in an OTU.

The two descendant lineages or clades from a single common ancestor are known as **sister groups** or **sister taxa**. For example, in Figure 2.6E and F, sister group pairs are: (1) *X. lutea* and *X. alba*; (2) *X. nigra* and *X. purpurea*; (3) *X. nigra* + *X. purpurea* and *X. rubens*; and (4) *X. lutea* + *X. alba* and *X. nigra* + *X. purpurea* + *X. rubens*.

The converse of monophyly is paraphyly. A **paraphyletic group** is one that includes a common ancestor and some, *but not all*, known descendants of that ancestor. For example, in Figure 2.6E, a group including ancestor *Q* plus descendants *X. lutea*, *X. alba*, and *X. rubens* alone is paraphyletic because this group has omitted two taxa (*X. purpurea* and *X. nigra*), which are also descendants of common ancestor *Q*. Paraphyletic groups have usually been designated by the absence of an apomorphy; in the example of Figure 2.6E, *X. lutea*, *X. alba*, and *X. rubens* may have originally been grouped by their lack of the derived condition (two stamens, character 5) that links *X. purpurea* and *X. nigra.* Similarly, a **polyphyletic group** is one containing two or more common ancestors. For example, in Figure 2.6E, a group containing *X. alba* and *X. purpurea* alone could be interpreted as polyphyletic because these two taxa do not share a common ancestor apart from that shared by the other taxa, *X. lutea*, *X. nigra,* and *X. rubens*. Polyphyletic groups have typically been defined based on convergences; in the example of Figure 2.6E, *X. lutea* and *X. purpurea* share a feature (spiny pollen of character 6), which has been determined from the analysis to be convergent in the two taxa. ("Paraphyletic" and "polyphyletic" as designates for a group may intergrade; the term **non-monophyletic** may be used to refer to either.)

Non-monophyletic groups do not accurately portray evolutionary history and should be abandoned in formal classification systems (see **Phylogenetic Classification**, page 41). Their use in comparative studies of character evolution, evolutionary processes, ecology, or biogeography will likely bias the results. A good example of a paraphyletic group is the traditionally defined "dicots" (Dicotyledonae). Because all recent analyses

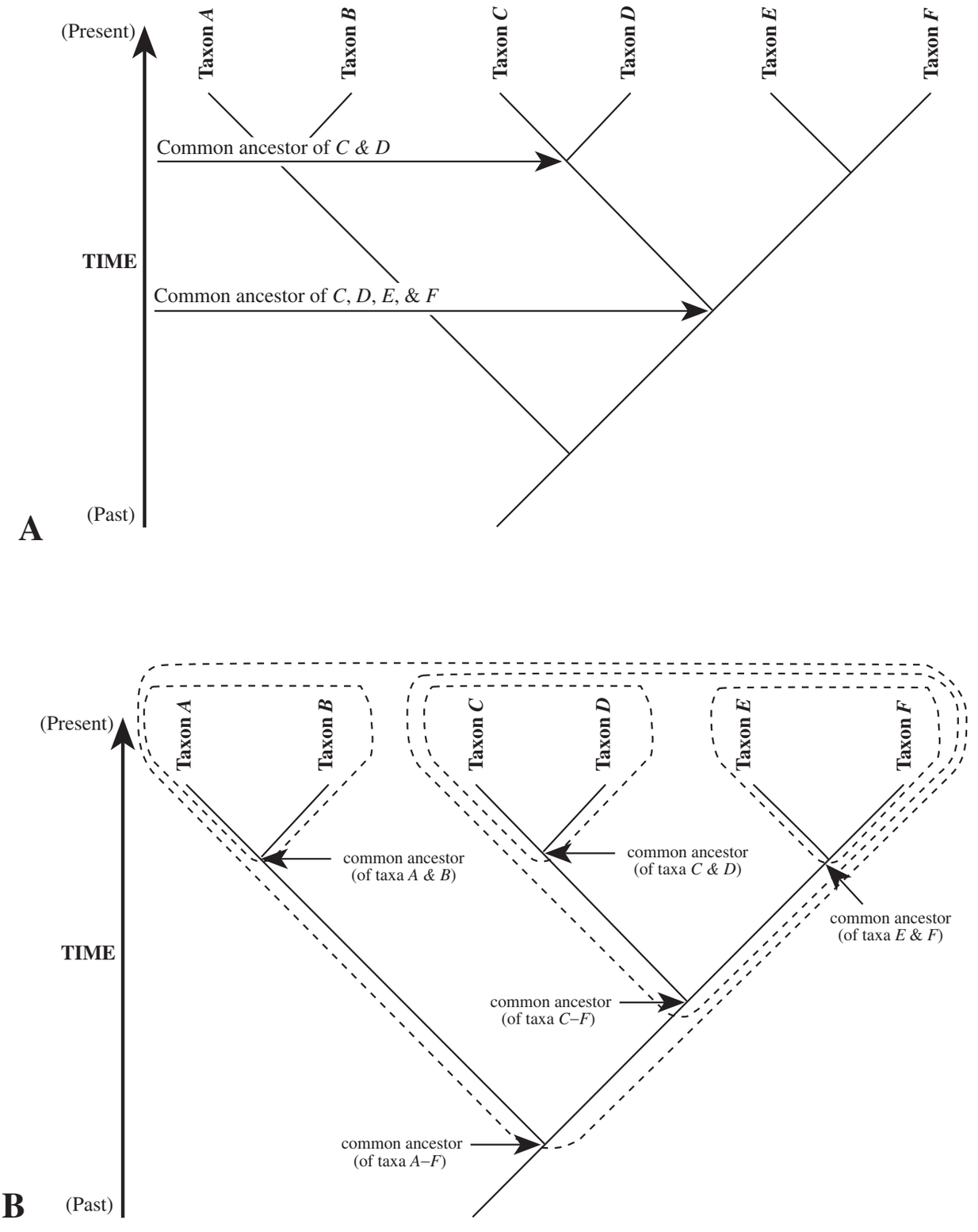


FIGURE 2.7 **A.** Hypothetical cladogram, illustrating recency of common ancestry. **B.** Cladogram of A with all monophyletic groups circled.

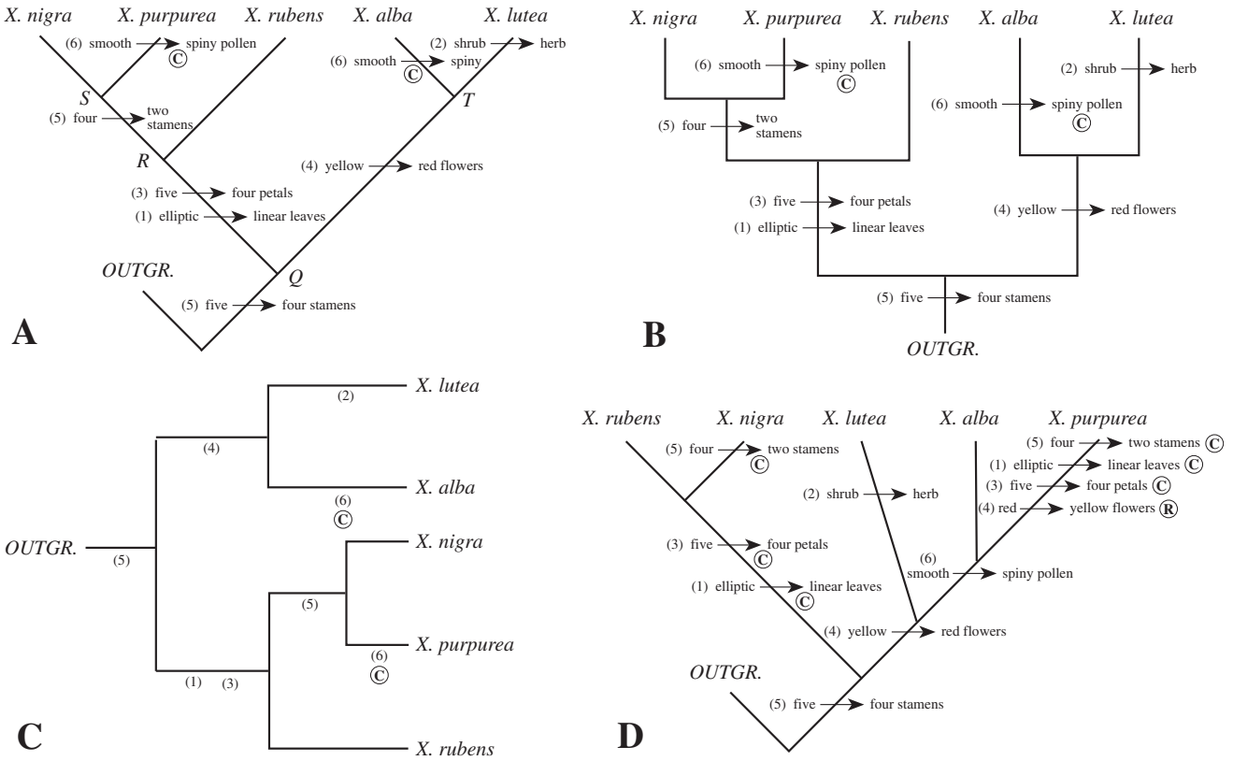


FIGURE 2.8 **A–C.** Most parsimonious cladogram of Fig. 2.6E. **A.** Cladogram with diagonal lines, but lineages rotated about common ancestor. **B.** Cladogram portrayed with perpendicular lines. **C.** As in B but rotated 90°. **D.** Alternative cladogram for the data set of Fig. 2.6A, showing a different relationship among the five species of genus *Xid*, requiring 11 character state changes, three more than the most parsimonious cladogram in Fig. 2.6E. Convergent and reversal homoplasies are denoted by circled **C** and **R**, respectively.

show that some members of the dicots are more closely related to, e.g., monocots (Monocotyledonae) than they are to other dicots, the dicots are paraphyletic (see Chapter 7) and should no longer be recognized as a taxonomic group.

PARSIMONY ANALYSIS

In constructing a cladogram, a single branching pattern may be selected from among many, many possibilities. The number of possible "rooted" dichotomously branching cladograms increases dramatically with a corresponding increase in the number of taxa. For two taxa, there is only one rooted cladogram (Figure 2.9A); for three taxa, there are three rooted, dichotomously branched trees (Figure 2.9B); and for four taxa, 15 rooted, dichotomously branched cladograms are possible (Figure 2.9C). (See later discussion for explanation of rooted versus unrooted trees.) The formula for the number of rooted trees is $\prod (2i-1)$, with $\prod$ being the product of all the factors $(2i-1)$ from $i = 1$ to $i = n-1$, where n is the number of OTUs. For a cladistic analysis involving 54 OTUs, the number of possible dichotomously branching trees is 3×10^{84} (which is greater than the number of atoms in the universe!). The number of trees is even greater when the additional possibilities of reticulation or polytomies are taken into account (see later discussion).

Because there are generally many possible trees for any given data set, one of the major methods of reconstructing phylogenetic relationships is known as the **principle of parsimony** or **parsimony analysis**. The principle of parsimony states that of the numerous possible cladograms for a given group of OTUs, the one (or more) exhibiting the fewest number of evolutionary steps is accepted as being the best estimate of phylogeny. (Note that there may be two or more cladograms that are equally most parsimonious.) The principle of parsimony is actually a specific example of a general tenet of science known as Ockham's Razor ("*Entia non sunt multiplicanda praeter necessitatem*"), which states that given two or more competing hypotheses, each of which can explain the facts, the simplest one is accepted. The rationale for parsimony analysis is that the simplest explanation minimizes the number of ad hoc hypotheses, i.e., hypotheses for which there is no direct evidence. In other words, of all possible cladograms for a given group of taxa, the one (or more)

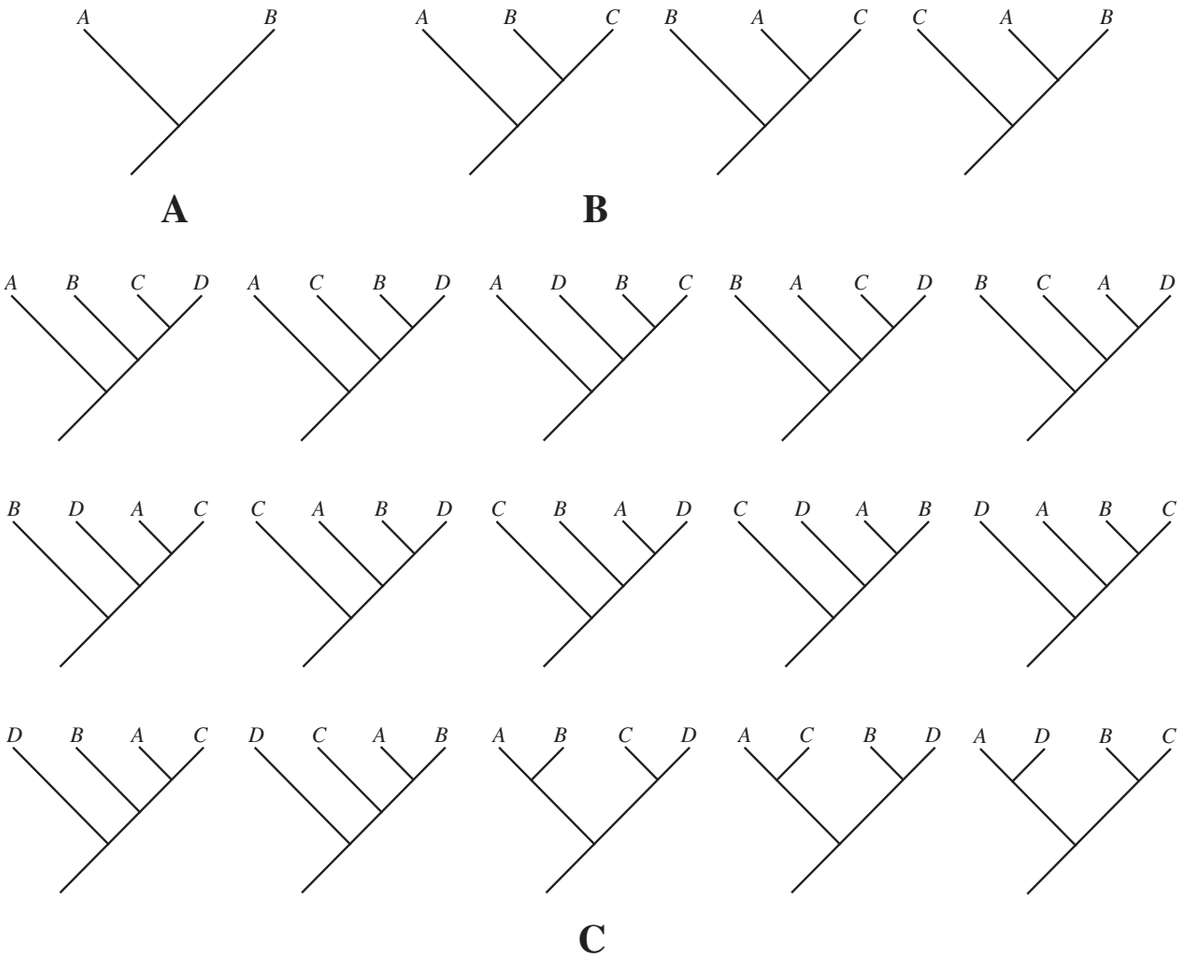


FIGURE 2.9 All possible rooted, dichotomously branched cladograms for a group consisting of the following. **A.** Two taxa (*A* and *B*). **B.** Three taxa (*A*, *B*, and *C*). **C.** Four taxa (*A*, *B*, *C*, and *D*).

implying the fewest number of character state changes is accepted. A consequence of minimizing the total number of character state changes is also to minimize the number of homoplasious reversals or convergences. The principle of parsimony is a valid working hypothesis because it minimizes uncorroborated hypotheses, thus assuming no additional evolutionary events for which there is no evidence.

Parsimony analysis can be illustrated as follows. For the example data set of Figure 2.6A, which includes five taxa (plus an outgroup), there are actually 105 *possible* dichotomously branching, rooted cladograms; the cladogram in Figures 2.6E and 2.8A (having a total of eight character state changes) is only one of these. One of the other 104 alternative cladistic hypotheses is illustrated in Figure 2.8D. Note, however, that for this cladogram, there are a minimum of 11 character state changes (including three pairs of convergent evolutionary events and one reversal). Thus, of all the possible cladograms for the data set of Figure 2.6A, the *one* shown in Figures 2.6E and 2.8A is the shortest, containing the fewest number of evolutionary steps, and would be accepted as the best estimate of phylogeny by parsimony analysis.

Various computer programs (algorithms) are used to determine the most parsimonious cladogram from a given character × taxon matrix. (See **Cladistic Computer Programs**, page 52.)

UNROOTED TREES

In contrast to a cladogram, a method for the representation of relative character state changes between taxa is the unrooted tree, sometimes called a "network." Unrooted trees are constructed by grouping taxa from a matrix in which polarity is not indicated (in which no hypothetical ancestor or outgroup

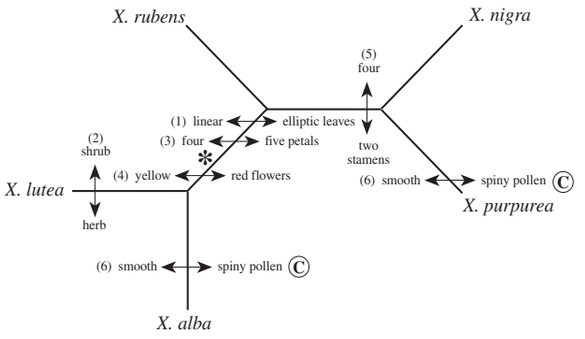


FIGURE 2.10 Unrooted tree for the data set of Fig. 2.6A (minus the *Outgroup* taxon). Direction of evolutionary change is not indicated and monophyletic groups cannot be defined. The "*" indicates the point of rooting that yields the tree of Fig. 2.6E.

is designated), perhaps because the polarity of one or more characters cannot be ascertained. Because no assumptions of polarity are made, no evolutionary hypotheses are implicit in an unrooted tree. Figure 2.10 illustrates the unrooted tree for the data set of Figure 2.6A,B. Note that monophyletic groups cannot be recognized in unrooted trees because relative ancestry (and therefore an outgroup) is not indicated. The character state changes noted on the unrooted tree simply denote evolutionary changes when going from one group of taxa to another, without reference to direction of change. After an unrooted tree is constructed, it may be rooted and portrayed as a cladogram. If the relative ancestry of one or more characters can be established, a point on the network may be designated as most ancestral, forming the root of the cladogram. For example, if the unrooted tree of Figure 2.10 is rooted (at * in the figure), the result is the tree of Figure 2.6E. However, rooting is effectively done by simply including one or more outgroup(s) in the analysis and placing these outgroups to denote the base (the "root") of the tree.

POLYTOMY

Occasionally, the relationships among taxa cannot be resolved. A **polytomy** (also called a polychotomy) is a branching diagram in which the lineages of three or more taxa arise from a single hypothetical ancestor. Polytomies arise either because data are incomplete or missing, because of conflicting data, or because three or more of the taxa were actually derived from a single ancestral species. (In addition, polytomies are often found in consensus trees; see later discussion.)

In the case of a polytomy arising via missing data, there are no derived character states identifying the monophyly of any two taxa among the group. For example, from the **character × taxon matrix** of Figure 2.11A, the relationships among taxa *W*, *X*, and *Y* cannot be resolved; synapomorphies link none of the taxon pairs. Thus, *W*, *X*, and *Y* are grouped as a polytomy in the most parsimonious cladogram (Figure 2.11B).

Polytomies often arise because of conflicting data in a *consensus tree* (see later discussion). Lastly, another possible reason for the occurrence of a polytomy is that three or more taxa actually diverged from a single ancestral species, or diverged dichotomously but within such a short time period that no synapomorphic evolutionary event links any two of the taxa as a monophyletic group. (See Chapter 19.) The occurrence of polytomies in phylogenetic analysis may be indicative of an interesting evolutionary history, warranting further research.

RETICULATION

The methodology of phylogenetic systematics generally presumes the dichotomous or polytomous splitting of taxa, representing putative ancestral speciation events. However, another possibility in the evolution of plants is **reticulation**, the hybridization of two previously divergent taxa forming a new lineage. A reticulation event between two ancestral taxa (*E* and *F*) is exemplified in Figure 2.11D (assuming transmission of derived states of characters 1-3 to hybrid taxon *G*), resulting in the hybrid ancestral taxon *G*, which is the immediate ancestor of extant taxon *X*. Most standard phylogenetic analyses do not consider reticulation and would yield an incorrect cladogram if such a process had occurred. For example, the character × taxon matrix of Figure 2.11C is perfectly compatible with the reticulate cladogram of Figure 2.11D. However, the methods of phylogenetic systematics would construct the most parsimonious *dichotomously branching* cladogram of Figure 2.11E or 2.11F, which show homoplasy and require one additional character state change than Figure 2.11D.

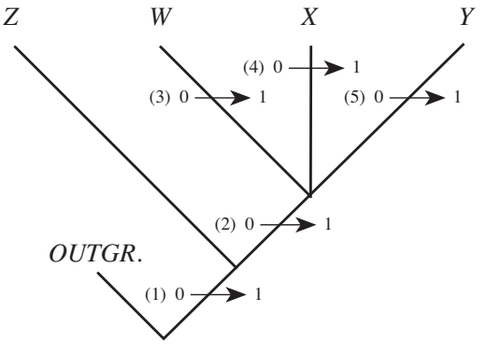
Reticulation among a group of taxa should always be treated as a possibility. Data, such as chromosome analysis, may provide compelling evidence for past hybridization among the most recent common ancestors of extant taxa. A good example of this is the evolution of durum and bread wheat (*Triticum* spp.) via past hybridization and polyploidy (Figure 2.11G).

TAXON SELECTION AND POLYMORPHIC CHARACTERS

As alluded to earlier, the initial selection of taxa to be studied may introduce bias in a phylogenetic analysis. Prior to a phylogenetic analysis, each of the smallest unit taxa under study (OTUs) *and* the group as a whole must be assessed for monophyly prior to the analysis. Monophyly is ascertained by the recognition of one or more unique, shared derived character states that argue for most recent common ancestry of all and

	1	2	3	4	5
W	1	1	1	0	0
X	1	1	0	1	0
Y	1	1	0	0	1
Z	1	0	0	0	0
OUTGR.	0	0	0	0	0

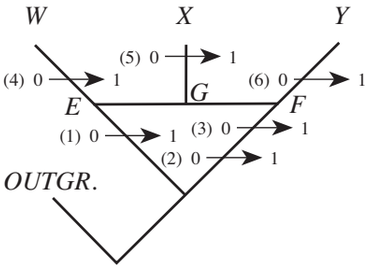
A



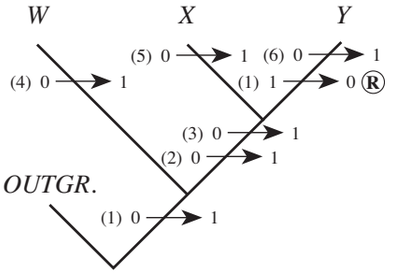
B

	1	2	3	4	5	6
W	1	0	0	1	0	0
X	1	1	1	0	1	0
Y	0	1	1	0	0	1
OUTGR.	0	0	0	0	0	0

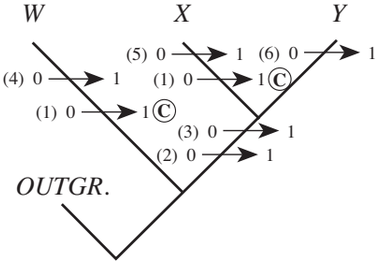
C



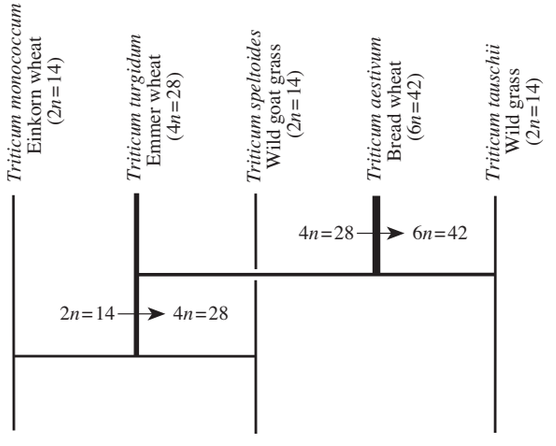
D



E

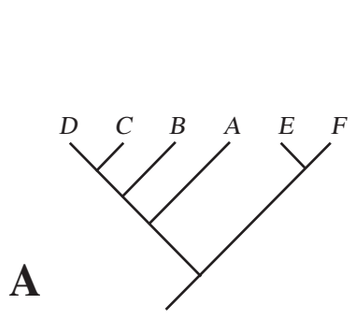


F

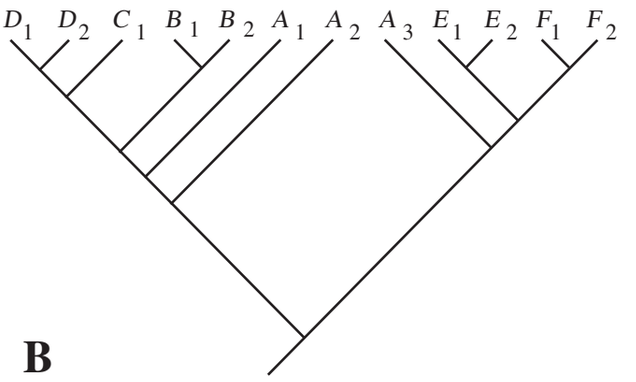


G

FIGURE 2.11 **A.** Hypothetical data set. **B.** Resultant tree from data set at A. Note polytomy of lineages to *W*, *X*, and *Y*. **C.** Hypothetical data set. **D.** Cladogram exhibiting reticulation that is compatible with data set at C. **E,F.** Dichotomously branching cladograms arising from data set at C, showing two alternative distributions of character state changes. **G.** Evolution of wheat via ancestral hybridization and polyploidy.



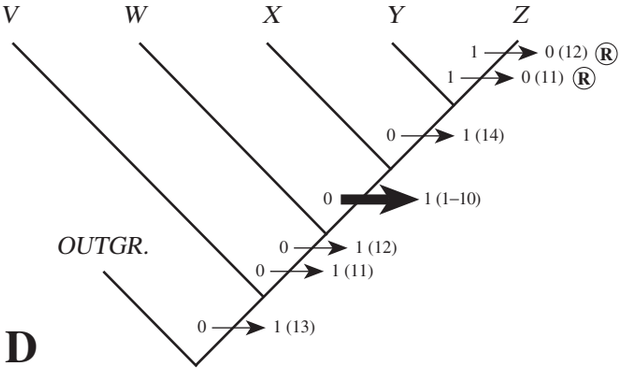
A



B

	1–10	11	12	13	14
V	0	0	0	1	0
W	0	1	1	1	0
X	1	1	1	1	0
Y	1	1	1	1	1
Z	1	0	0	1	1
OUTGR.	0	0	0	0	0

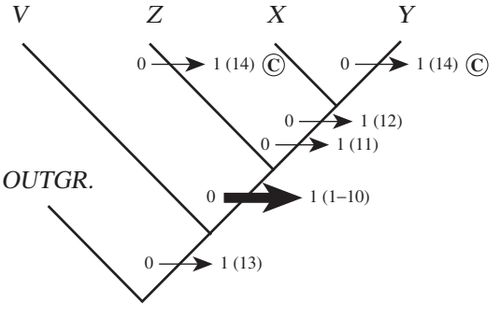
C



D

	1–10	11	12	13	14
V	0	0	0	1	0
X	1	1	1	1	0
Y	1	1	1	1	1
Z	1	0	0	1	1
OUTGR.	0	0	0	0	0

E



F

FIGURE 2.12 **A.** Cladogram of hypothetical genera *A–F*. **B.** Cladogram of the species of genera *A–F*. Note that genus *A* is not monophyletic. **C.** Character × taxon matrix for taxa *V–Z* plus *OUTGROUP*. **D.** Most parsimonious cladogram for taxa *V*, *W*, *X*, *Y*, and *Z*. **E.** Character × taxon matrix for same taxa, minus taxon *W*, omitted because it is not considered as part of the ingroup. **F.** Most parsimonious cladogram for taxa *V*, *X*, *Y*, and *Z*. Note different branching pattern for taxa *X*, *Y*, and *Z*.

only all members of the taxon in question. If such an apomorphy cannot be identified, any relationships denoted from the phylogenetic analysis may be in doubt. For example, in a cladistic analysis of several angiosperm genera (Figure 2.12A), only if each of the unit taxa (genera in this case) is monophyletic will the resultant cladogram be unbiased. If, however, genus *A* is not monophyletic, then it may be possible for some species of genus *A* to be more closely related to (i.e., have more recent common ancestry with) a species of another genus than to the other species of genus *A* (e.g., Figure 2.12B). Therefore, if any doubt exists as to the monophyly of component taxa to be analyzed, the taxa in question should be subdivided until the monophyly of these subtaxa is reasonably certain. If this is not possible, an exemplar species (selected as representative of a

higher taxon and assumed to be monophyletic) may be chosen for a first approximation of relationships.

Related to the requirement of OTU monophyly is the problem of **polymorphic** characters, i.e., those that have variable character state values *within* an OTU. If an OTU for which monophyly has been established is polymorphic for a given character, then it may be subdivided into smaller taxonomic groups until each of these groups is monomorphic (i.e., invariable) for the character. If an OTU at the level of species is polymorphic, it is generally listed as such in computer algorithms.

If the ingroup as a whole is not monophyletic, the effect is identical to excluding taxa from the analysis, which could give erroneous results under certain conditions. For example, the most parsimonious cladogram constructed from the data matrix of Figure 2.12C is that of Figure 2.12D. However, if taxon *W* is inadvertently omitted from the ingroup (which is now not monophyletic; Figure 2.12E), then a different, most parsimonious cladogram topology may result for taxa *X*, *Y*, and *Z* (Figure 2.12F). The question of monophyly may be a serious problem for traditionally recognized taxa that were generally not defined by demonstrable apomorphies.

OUTGROUP COMPARISON

As mentioned earlier in the discussion on character analysis, knowledge of character polarity is necessary to recognize shared derived character states that define monophyletic taxa. The only valid criterion for ascertaining polarity is outgroup comparison. An **outgroup** is a taxon that is not a member of the study group under investigation (the **ingroup**). Outgroup comparison entails character assessment of the *closest* outgroups to the ingroup. Those character states possessed by the closest outgroups are considered to be ancestral; states present in the ingroup, but not occurring in the nearest outgroups, are derived.

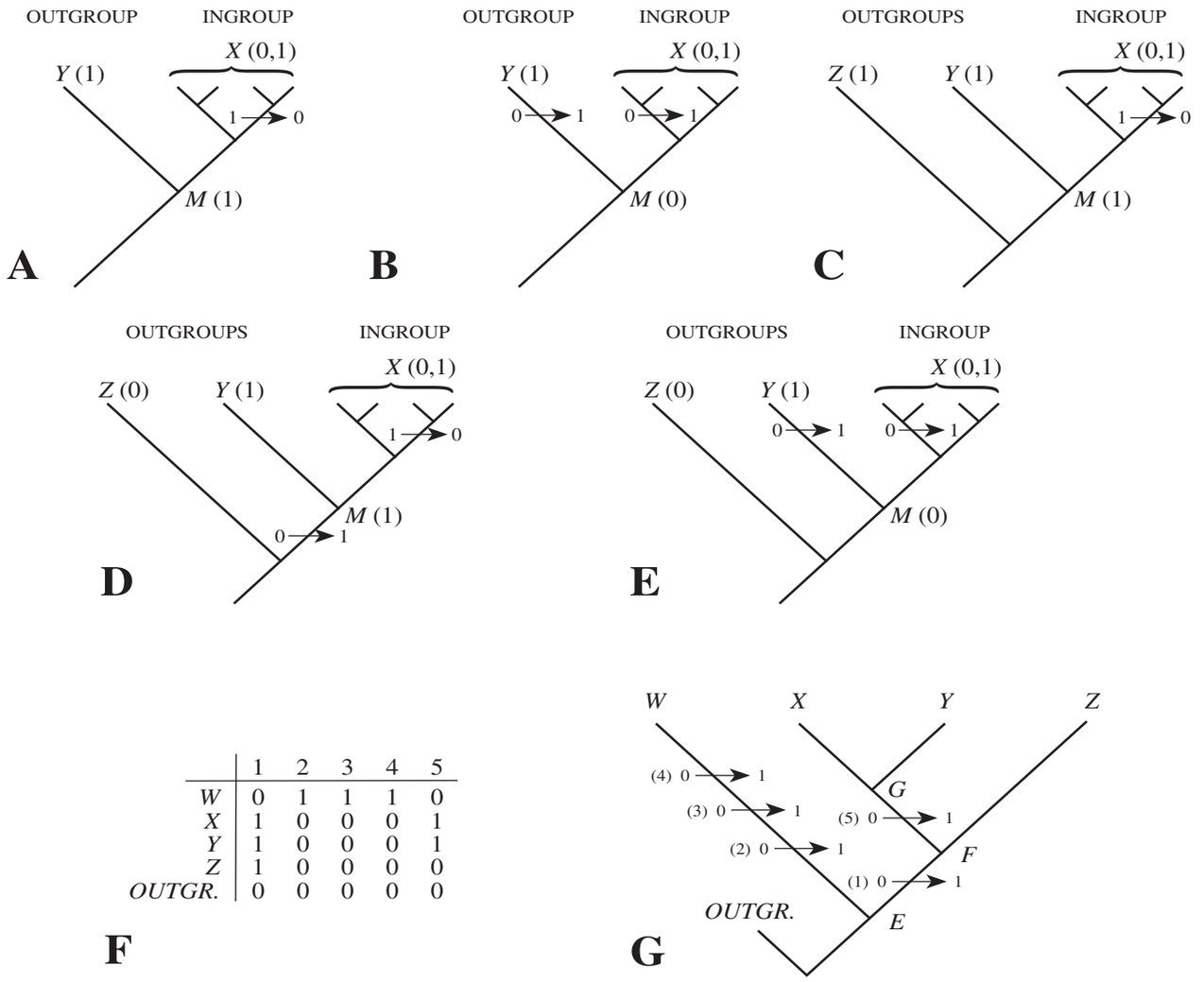
The rationale for outgroup comparison is founded in the principle of parsimony. For example, given some monophyletic ingroup *X* (Figure 2.13A), members of which possess either state 0 or 1 of a character, and given that taxon *Y* (nearest outgroup to *X*) possesses only character state 1, then the most parsimonious solution (requiring a single character change: $1 \Rightarrow 0$) is that state 1 is ancestral and present in the common ancestor *M* (the "outgroup node"); character state 0 is derived within taxon *X* (Figure 2.13A). The alternative, that state 0 is ancestral, requires at least two character state changes (Figure 2.13B). Verification is made by considering an additional outgroup (e.g., taxon *Z* in Figure 2.13C). If this next outgroup possesses only character state 1, then the ancestral status of state 1 for taxon *Y* is substantiated (Figure 2.13C). If, however, outgroup *Z* contains only character state 0, then it is equally parsimonious to assume that state 1 is ancestral to ingroup *X* (Figure 2.13D) versus derived within ingroup *X* (Figure 2.13E). In this case, consideration of additional outgroups may resolve polarity. The major problem with outgroup comparison is that the cladistic relationships of outgroup taxa may be unknown; in such a case, all reasonably close outgroups (in all possible combinations) may be tested. In practice, prior studies at a higher taxonomic level are often used to establish near outgroups for a phylogenetic analysis.

ANCESTRAL VERSUS DERIVED CHARACTERS

A common point of confusion is seen in the use of the terms *ancestral* (plesiomorphic or primitive) and *derived* (apomorphic or advanced). It is advisable that these terms be limited to the description of characters (not taxa) and then only relative to monophyletic groups. For example, in the cladogram of Figure 2.13G (constructed from the matrix of Figure 2.13F), state 1 of character 1 is derived *within* the group including *W*, *X*, *Y*, and *Z* (i.e., state 1 is absent in common ancestor *E*), but it is ancestral with regard to the monophyletic group *X*, *Y*, *Z* (i.e., state 1 is present in *F*, the common ancestor of *X*, *Y*, and *Z*). The use of the terms ancestral and derived to describe *taxa* should be avoided to prevent ambiguity. For example, from Figure 2.13G, it might be asked which taxon is most "primitive"? Confusion is avoided by describing, e.g., taxon *W* as phylogenetically most "basal" (or "earliest diverging") and, e.g., taxon *Z* as possessing the fewest number of observed apomorphic states (relative to a common ancestor, such as ancestral taxon *E*).

CONSENSUS TREES

In practice, most cladistic analyses yield numerous cladograms that are equally most parsimonious. Rather than view and discuss each of these cladograms, it is usually convenient to visualize the one tree that is compatible with all equally most parsimonious trees. A **consensus tree** is a cladogram derived by combining the features in common between two or more cladograms. There are several types of consensus trees. One of the most commonly portrayed is the **strict consensus tree**, which collapses differences in branching pattern between two or more cladograms to a polytomy. Thus, the two equally parsimonious cladograms of Figure 2.14A,B are collapsible to the strict consensus tree of Figure 2.14C. Another type of consensus tree is the **50% majority consensus tree**, in which only those clades that occur in 50% or more of a given set of trees are retained. Consensus trees may be valuable for



	1	2	3	4	5
W	0	1	1	1	0
X	1	0	0	0	1
Y	1	0	0	0	1
Z	1	0	0	0	0
OUTGR.	0	0	0	0	0

F

FIGURE 2.13 Determination of character state polarity using outgroup comparison. **A.** Most parsimonious assumption in which character state 1 is ancestral and present in ancestor *M*. **B.** Alternative, less parsimonious cladogram, in which state 0 is assumed to be ancestral. **C.** Verification of cladogram at A by addition of next outgroup *Z*, which also has state 1. **D,E.** Cladograms in which additional outgroup *Z* has state 0, showing that assumption of polarity is equivocal; ancestor *M* is equally likely to possess state 0 as opposed to state 1. **F.** Character × taxon matrix for taxa *W–Z* plus *OUTGROUP*. **G.** Most parsimonious cladogram of taxa *W*, *X*, *Y*, and *Z* and ancestors *E*, *F*, and *G*.

assessing those clades that are robust, i.e., have strong support (see **Cladogram Robustness**, page 39). Greater confidence may be given to such clades in terms of recognition of accepted and named monophyletic groupings.

LONG BRANCH ATTRACTION

Sometimes, e.g., with molecular sequence data, one or more taxa will have a very long branch, meaning that these taxa have a large number of autapomorphies relative to other taxa in the analysis (e.g., taxon *Z* of Figure 2.14D). This can be caused by unequal rates of evolution among the taxa examined or can be the by-product of the particular data used. Such a situation can result in "long branch attraction," in which taxa with relatively long branches tend to come out as close relatives of one another (or, if only one taxon has a long branch, its phylogenetic placement may easily shift from one analysis to another). Long branch attraction occurs because when relatively numerous state changes occur along lineages, random changes can begin to outweigh nonrandom, phylogenetically informative ones. The phylogenetic placement of a taxon with a long branch can be uncertain and can unduly influence the placement of other taxa.

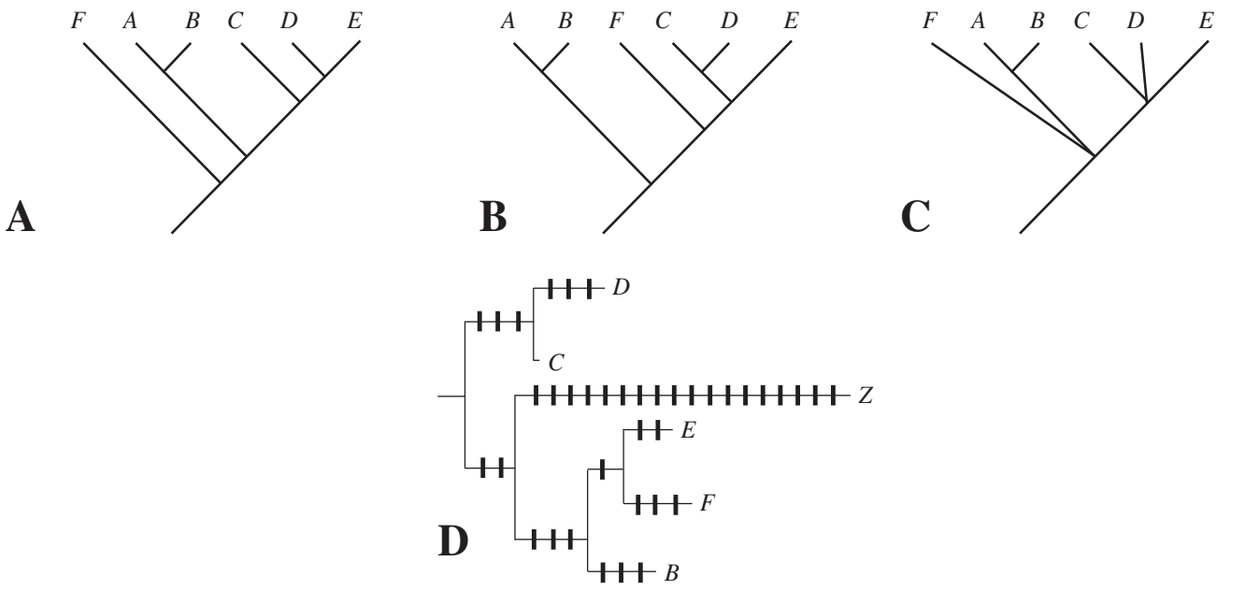


FIGURE 2.14 **A,B.** Two equally most parsimonious cladograms resulting from cladistic analysis. **C.** Strict consensus tree of cladograms at A and B. **D.** Cladogram illustrating taxon *Z* with very long branch.

Taxa with long branches may need to be analyzed using a different data set. They are sometimes left out of an analysis to see what the effect is on cladogram robustness (see later discussion).

MAXIMUM LIKELIHOOD

The principle of parsimony can be viewed as evaluating all alternative trees (or as many subsets as feasible), calculating the length of those trees, and selecting those trees that are shortest, i.e., require the minimum number of character state changes under the set of conditions (character coding) specified. Another method of phylogenetic inference is termed **maximum likelihood**. Maximum likelihood, like parsimony methods, also evaluates alternative trees (hypotheses of relationship), but considers the *probability*, based on some selected *model of evolution*, that each tree explains the data. That tree which has the highest probability of explaining the data is preferred over trees having a lower probability. The appropriate model of evolution used is typically based on the data of the current analysis, but may be based on other data sets.

Maximum likelihood is used in practice for molecular sequence data, although morphological data or a combination of the two can be used. Figure 2.15A shows a simple molecular data set of three characters (three nucleotide sites of some gene or gene region). In this example, there are three possible trees, shown as unrooted in Figure 2.15B and rooted at taxon *Z* (assuming this is the outgroup) in Figure 2.15C. Maximum likelihood evaluates each tree and calculates, for each character, the total *probability* that each node of the tree possesses a given nucleotide. (See Chapter 14 for information on molecular data.)

In this same example, Figure 2.16A shows the actual probability of a change from one base to another, i.e., via nucleotide substitutions. Thus, a change from **A** (adenine) to **C** (cytosine) has a probability of 0.1 (10%), that from an **A** to **G** has a probability of 0.2 (20%), etc. Consider the first unrooted tree of Figure 2.15B for the first character (site 43) of Figure 2.15A. Nucleotide bases (A, C, G, or T) are substituted for taxon names (W–Z), and each of the two internal nodes (ultimately corresponding to hypothetical common ancestors) are arbitrarily assumed to possess **A** (adenine), shown in Figure 2.16B. The overall probability for this nucleotide combination on this particular tree is the starting probability of any particular nucleotide (0.25 in this example, with the assumption that nucleotide bases are in equal frequency, each being 25% given there are 4 bases) × the probability of going from an **A** to a **C** (= 0.1) × the probability of going from an **A** to a **G** (= 0.2) × the probability of going from an **A** to an **A** (= 0.60), and so on; the total probability for this tree topology and base combination is P_1=0.00003 (Figure 2.16B). Now, the total probabilities for all 16 possible combinations of nucleotide bases at the internal nodes is seen in Figure 2.16C (P_1 ... P_{16}). The *likelihood* score for this tree and site (character 1) is calculated by adding all of these individual probabilities (P_1 + P_{2+} ... P_{16}) = 0.0026 in this example (Figure 2.16C).

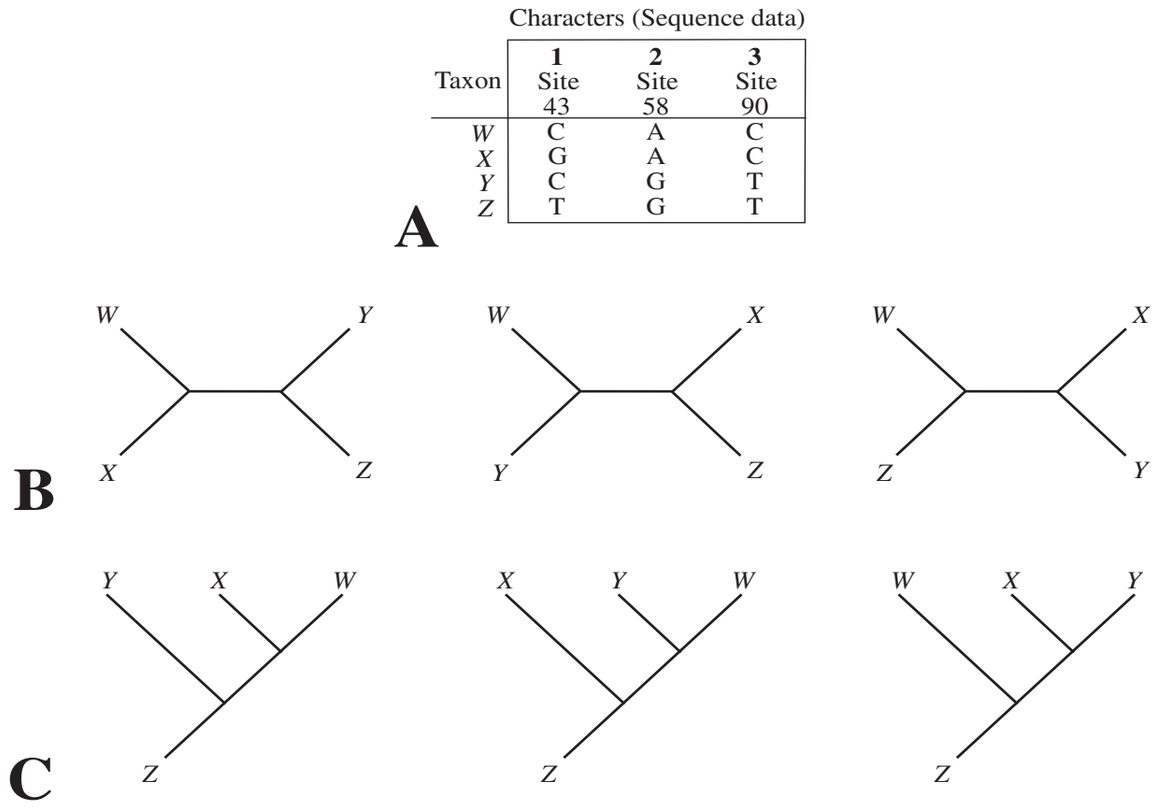


FIGURE 2.15 Maximum likelihood. **A.** Example character × taxon matrix (three nucleotide base sites shown) for taxa *W–Z*. **B.** The three possible unrooted trees for taxa *W–Z*. **C.** Three possible rooted trees for taxa, with *Z* set as the root.

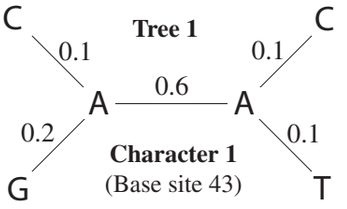
Likelihoods scores for each of the other two characters (sites 58 and 90 in Figure 2.15A) of this first tree topology are then calculated (Figure 2.16D,E). The total likelihood for this tree topology is obtained by multiplying the likelihood scores of each of the three characters. (Normally, the negative natural log of each probability is calculated and these added together, due to the often small numbers obtained for L.) In the example of Figure 2.16, $L = 0.0026 \times 0.01132 \times 0.01132 = 3.332 \times 10^{-7}$, or -LnL = -(Ln 0.0026 + Ln 0.1132 + Ln 0.1132) = 14.915. It turns out that the first tree represented [(W,X)(Y,Z)] has the highest likelihood (L) of the three possible unrooted trees and would be accepted over the other two.

Maximum likelihood uses a DNA substitution **model** to determine the probabilities of going from one nucleotide to another. These models and algorithms are complicated, but the very basics are important to grasp. One that is commonly used and serves as the basis for other specific models is the **general time-reversible** model (**GTR**), in which a change from one base to another (e.g., A to C) is equivalent to the reverse (e.g., C to A). As seen in Figure 2.17, the GTR model is based on substitution probabilities that are influenced by the *rate parameter* (e.g., μa), which is the product of the mean instantaneous substitution rate (μ) and the relative rate parameters (*a*, *b*, ... *f*), those for each substitution type (e.g., A to C) and the *frequency parameters* (π_A, π_C, π_G, and π_T), which are the frequencies of the nucleotide bases A, C, G, and T. (Other assumptions are made in this model; see Hillis et al., 1996.) Specific models may be derived from the GTR model. For example, if the frequency parameters are equivalent ($\pi_A = \pi_C = \pi_G = \pi_T = 0.25$, given there are 4 bases), and if all substitutions occur at the same rate (a = b = c = d = e = f = 1), then the Jukes-Cantor (JC) model is obtained (Figure 2.17B). If the frequency parameters are equal, but substitutions occur at different rates, such that all transition rates (A⇔G and C⇔T) are equivalent but potentially different from all transversions (A⇔C, A⇔T, G⇔C, and G⇔T), then Kimura's two parameter model (K2P) is obtained (Figure 2.17C). Other models might take into account, e.g., the codon position of a base. The model that is used in a maximum likelihood analysis is calculated from the actual sequence data using a computer algorithm. For example, if significantly more transitions occur than transversions, the K2P model might be selected. (See WEB SITE, Phylogeny Programs,

A

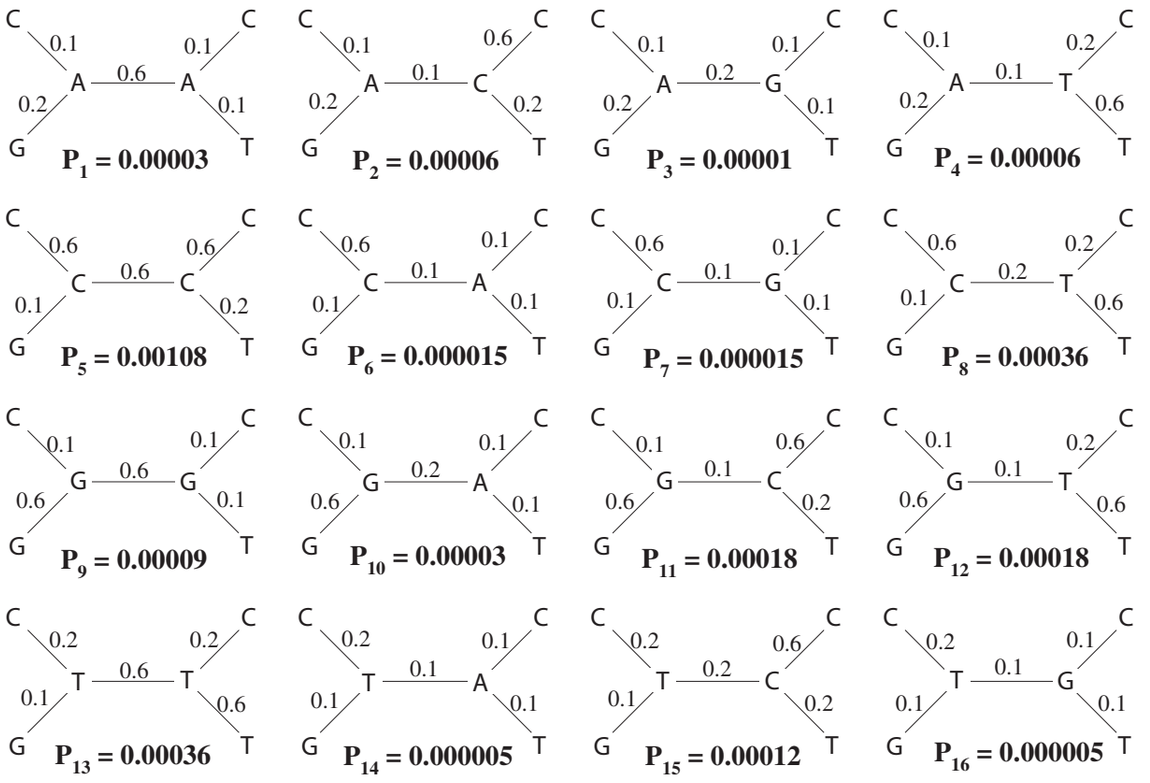
	A	C	G	T
A	0.6	0.1	0.2	0.1
C	0.1	0.6	0.1	0.2
G	0.2	0.1	0.6	0.1
T	0.1	0.2	0.1	0.6

B



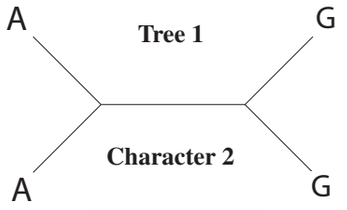
Probability$_1$ (P_1) = (0.25)•(0.1)•(0.2)•(0.6)•(0.1)•(0.1) = 0.00003

C



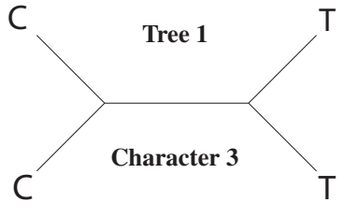
$L_1 = P_1 + P_2$ ..... + P_{16} = (0.00003) + (0.00006) ..... + (0.000005) = 0.0026

D



L_2 = 0.01132

E



L_3 = 0.01132

FIGURE 2.16 Maximum likelihood. **A.** Matrix showing the probabilities of a change from one base to another. **B.** Tree 1 (one of three possible unrooted trees from Figure 2.15B), with nucleotide bases of character 1 substituted for terminal taxa and with internal nodes arbitarily set to A (adenine). Note calculation of probability of site 1 (P_1). **C.** The sixteen possibilities for internal node bases for tree 1, with probabilities calculated. **D,E.** Representation of nucleotides at terminal taxa in tree 1 for characters 2 (D) and 3 (E), with likelihood calculations below.

A

	A	C	G	T
A	$-\mu(a\pi_C + b\pi_G + c\pi_T)$	$\mu a\pi_C$	$\mu b\pi_G$	$\mu c\pi_T$
C	$\mu a\pi_A$	$-\mu(a\pi_A + d\pi_G + e\pi_T)$	$\mu d\pi_G$	$\mu e\pi_T$
G	$\mu b\pi_A$	$\mu d\pi_C$	$-\mu(b\pi_A + d\pi_C + f\pi_T)$	$\mu f\pi_T$
T	$\mu c\pi_A$	$\mu e\pi_C$	$\mu f\pi_G$	$-\mu(c\pi_A + d\pi_C + e\pi_G)$

B

	A	C	G	T
A	$-3/4\mu$	$1/4\mu$	$1/4\mu$	$1/4\mu$
C	$1/4\mu$	$-3/4\mu$	$1/4\mu$	$1/4\mu$
G	$1/4\mu$	$1/4\mu$	$-3/4\mu$	$1/4\mu$
T	$1/4\mu$	$1/4\mu$	$1/4\mu$	$-3/4\mu$

C

	A	C	G	T
A	$-1/4\mu(\kappa+2)$	$1/4\mu$	$1/4\mu\kappa$	$1/4\mu$
C	$1/4\mu$	$-1/4\mu(\kappa+2)$	$1/4\mu$	$1/4\mu\kappa$
G	$1/4\mu\kappa$	$1/4\mu$	$-1/4\mu(\kappa+2)$	$1/4\mu$
T	$1/4\mu$	$1/4\mu\kappa$	$1/4\mu$	$-1/4\mu(\kappa+2)$

FIGURE 2.17 Models of base substitution. **A.** General time reversable model, in which probabilities of change from one base to another are a function of mean instantaneous base substitution rate (μ), relative rate parameters (a,b,c,d,e,f), and base frequencies (π_A, π_C, π_G, π_T). **B.** Jukes-Cantor (**JC**) model, in which substitution rates are the same. **C.** Kimura's two-parameter model (**K2P**), in which base frequencies are the same but transitions (in red) and transversions (in blue) occur at different rates.

page 52 for a listing of phylogeny computer programs, including those determining the model from a data set.)

Maximum likelihood methods have an advantage over parsimony in that the estimation of the pattern of evolutionary history can take into account probabilities of character state changes from a precise evolutionary model, one that is based and evaluated from the data at hand. Maximum likelihood methods also help eliminate the problem of long branch attraction (discussed earlier), as the probabilities of base change from one node to another are influenced by the length of that branch. (Generally, as the length of a branch increases, the probabilities of state changes along that branch decrease.) A disadvantage of maximum likelihood has been that both the analyses and calculation of confidence measures (usually bootstrap calculations; see **Cladogram Robustness**, page 39) have been very "computer-intensive" and generally not feasible for large data sets; however, new computer programs have dramatically increased the calculation speeds. See Hillis et al. (1996) for more detailed information about maximum likelihood and models.

BAYESIAN ANALYSIS

Another more recent method of phylogenetic analysis is **Bayesian** inference (which is worth mentioning briefly here,

but see the references at the end of this chapter for a detailed understanding). This method is based on calculations of **posterior probability**, utilizing a probability formula devised by T. Bayes in 1763.

Bayesian inference calculates the posterior probability of the phylogeny, branch lengths, and various parameters of the data. In practice, the posterior probability of phylogenies is approximated by sampling trees from the posterior probability distribution, using algorithms known as the *Markov chain Monte Carlo* (MCMC) or the *Metropolis-coupled Markov chain Monte Carlo* (MCMCMC). The results of a Bayesian analysis yield the posterior probabilities for each of the branches of a given tree (derived from the 50% majority consensus tree of sampled trees). Bayesian inference is similar to maximum likelihood in that the same models of evolution can be used. In addition, Bayesian algorithms are relatively rapid, and the posterior probabilities that are generated for each clade are direct measures of robustness. (Generally, a Bayesian probability of 95% or greater is considered robust for a particular clade: see **Cladogram Robustness**, right.)

MEASURES OF HOMOPLASY

If significant homoplasy occurs in a cladistic analysis, the data might be viewed as less than reliable for reconstructing phylogeny. One measure of the relative amount of homoplasy in the cladogram is the consistency index. **Consistency index** (**CI**) is equal to the ratio m/s, where m is the *minimum* number of character state changes that must occur and s is the *actual* number of changes that occur. The minimum number of changes is that needed to account for a single transformation between all character states of all characters. For example, a three-state character transformation, $0 \Leftrightarrow 1 \Leftrightarrow 2$, requires a minimum of two steps; e.g., one possibility (of several) is the change $0 \Rightarrow 1$ (first step) and then $1 \Rightarrow 2$ (second step).

A consistency index close to 1 indicates little to no homoplasy; a CI close to 0 is indicative of considerable homoplasy. As an example, the character × taxon matrix of Figure 2.6A,B necessitates a minimum of seven changes; i.e., there must be at least seven character state transformations to explain the distribution of states in the taxa. The actual number of changes in the most parsimonious cladogram is eight because of homoplasy (Figure 2.6E). Thus, the CI for this cladogram is 7/8 = 0.875. The consistency index may be viewed as a gauge of confidence in the data to reconstruct phylogenetic relationships.

A consistency index may be calculated for individual characters as well. For example, relative to the most parsimonious cladogram of Figure 2.6E, the CI of all characters is equal to 1, except for character 6, which has a CI of 0.5 (because of two convergent character state changes).

Two other measures of homoplasy may be calculated: the **retention index** (**RI**) and the **rescaled consistency index** (**RC**). The retention index is calculated as the ratio $(g - s)/(g - m)$, where g is the maximum possible number of state changes that could occur on any conceivable tree. Thus, the retention index is influenced by the number of taxa in the study. The rescaled consistency index (RC) is equal to the product of the CI and RI. The RC is used most often in successive weighting; the rationale for its use is based on theoretical simulation studies.

CLADOGRAM ROBUSTNESS

It is very important to calculcate metrics of **robustness**, the confidence for which a tree or particular clade actually denotes true phylogenetic relationships. A common way to evaluate cladogram robustness is the **bootstrap**, which can be used in both parsimony and maximum likelihood inference methods. **Bootstrapping** is a method that reanalyzes the data of the original character × taxon matrix by selecting (resampling) characters *at random*, such that a given character can be selected more than once. The effect of this resampling is that some characters are given greater weight than others, but the total number of characters used is the same as that of the original matrix. (See example in Figure 2.18A,B.) This resampled data is then used to construct new trees. Many sequential bootstrapping analyses are generated (often 100 or more runs), and all most parsimonious or greatest likelihood trees are determined. From all of these trees, a 50% majority consensus tree is constructed; the percentages placed along each internode of the cladogram represent the percentage of the time (from the bootstrap runs) that a particular clade is maintained from all resampled runs (e.g., Figure 2.18D). A bootstrap value of 70% or more is generally considered a robustly supported node. The rationale for bootstrapping is that differential weighting by resampling of the original data will tend to produce the same clades if the data are "good," i.e., reflect the actual phylogeny and exhibit little homoplasy. One problem with the bootstrapping method is that it technically requires a random distribution of the data, with no character correlation. These criteria are almost never verified in a cladistic analysis. However, bootstrapping is still the most used method to evaluate tree robustness.

Another method of measuring cladogram robustness occasionally used is the so-called **jackknife** (or **jackknifing**), which is similar to the bootstrap but differs in that each randomly selected character may only be resampled once (not multiple times), and the resultant resampled data matrix is

A

	01	02	03	04	05	06	07	08	09	10	11	12
T	0	1	0	0	1	0	1	1	1	1	0	1
U	1	1	1	0	1	0	0	1	1	1	1	0
V	0	0	0	1	1	0	1	0	1	1	0	1
W	1	1	1	0	0	1	0	0	1	0	1	0
X	1	1	0	0	1	1	0	1	0	1	0	0
Y	1	1	1	0	0	0	0	1	1	1	1	1

B

	01	11	11	02	03	04	05	06	06	10	10	12
T	0	0	0	1	0	0	1	0	0	1	1	1
U	1	1	1	1	1	0	1	0	0	1	1	0
V	0	0	0	0	0	1	1	0	0	1	1	1
W	1	1	1	1	1	0	0	1	1	0	0	0
X	1	1	1	1	0	0	1	1	1	1	1	0
Y	1	1	1	1	1	0	0	0	0	1	1	1

C

	01	02	03	04	05	06	10	12
T	0	1	0	0	1	0	1	1
U	1	1	1	0	1	0	1	0
V	0	0	0	1	1	0	1	1
W	1	1	1	0	0	1	0	0
X	1	1	0	0	1	1	1	0
Y	1	1	1	0	0	0	1	1

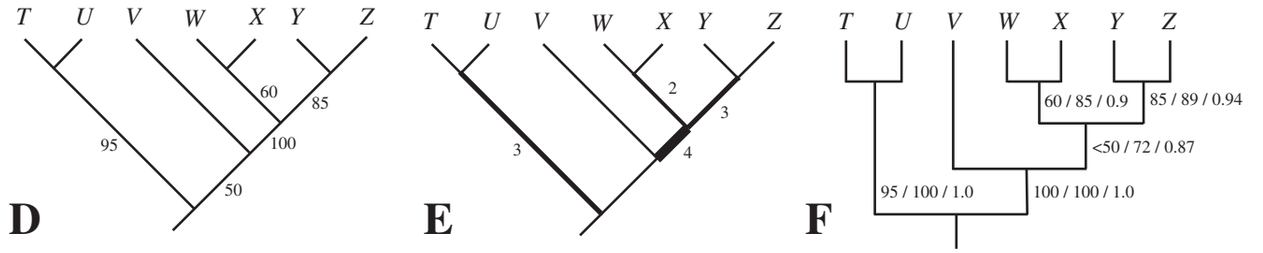


FIGURE 2.18 CLADOGRAM ROBUSTNESS. **A.** Data matrix of six taxa (T–Y) and 12 characters. **B.** Resampling of matrix, to be used in a bootstrap analysis. Note that the number of characters is the same and that some characters are repeated, some deleted. **C.** Resampling of matrix, to be used in a jackknife analysis. Note that no characters are repeated and the number of characters has been reduced. **D.** Cladogram showing parsimony bootstrap values at internodes with values > 50. **F.** Same cladogram as in E showing decay index values. (Increasing numbers correspond to increasing line thickness; internodes not numbered have a decay index of 1.) **F.** Same cladogram showing (left to right) parsimony bootstrap values, maximum likelihood bootstrap values, and Bayesian posterior probabilities.

smaller than the original. (See example in Figure 2.18A,C.) The resampled matrix is used to generate a tree or trees. This is repeated multiple times and, like the bootstrap, a 50% majority tree is created to generate jackknife values.

A second way to evaluate clade confidence is by measuring clade "decay." A **decay index** (also called "Bremer support") is a measure of how many extra steps are needed (beyond the number in the most parsimonious cladograms) before the original clade is no longer retained. Thus, if a given cladogram internode has a decay index of 4, then the monophyletic group arising from it is maintained even in cladograms that are four steps longer than the most parsimonious (e.g., Figure 2.18E). The greater the decay index value, the greater the "confidence" in a given clade.

Finally, Bayesian analysis provides a measure of robustness in calculating posterior probabilities for each of the clades generated. Any branch with a posterior probability of 95% or greater is statistically well-supported. However, this method has come under some scrutiny because it often generates particularly high values of support. In some analyses, parsimony bootstraps, maximum likelihood bootstraps, and Bayesian posterior probabilities may be indicated on the same cladogram, (e.g., a consensus tree), illustrating clade support from the three different analyses (see Figure 2.18F).

CLADOGRAM ANALYSIS

A typical cladistic analysis may involve the use of DNA sequence data from one or more genes plus the use of "morphological" (i.e., nonmolecular) data. (Tests may be used to evaluate the homogeneity or compatibility of phylogenetic information from different types of molecular data, e.g., from chloroplast versus nuclear genes.) Often, separate analyses are done for (1) each of the gene sets individually; (2) all molecular data combined; (3) morphological data alone; and (4) a combined analysis utilizing all available data—molecular and morphological. It has been demonstrated that utilizing the totality of data often results in the most robust cladogram. The strict consensus tree of this combined analysis generally represents the best estimate of phylogenetic relationships of the group studied.

From the most robust cladogram(s) derived from cladistic analyses, it is valuable to trace all character state changes. In addition, all monophyletic groupings should be evaluated in terms of their overall robustness (e.g., bootstrap support) and the specific apomorphies that link them together. Homoplasies (convergences or reversals) should also be noted. A homoplasy may represent an error in the initial analysis of that character that may warrant reconsideration of character state definition,

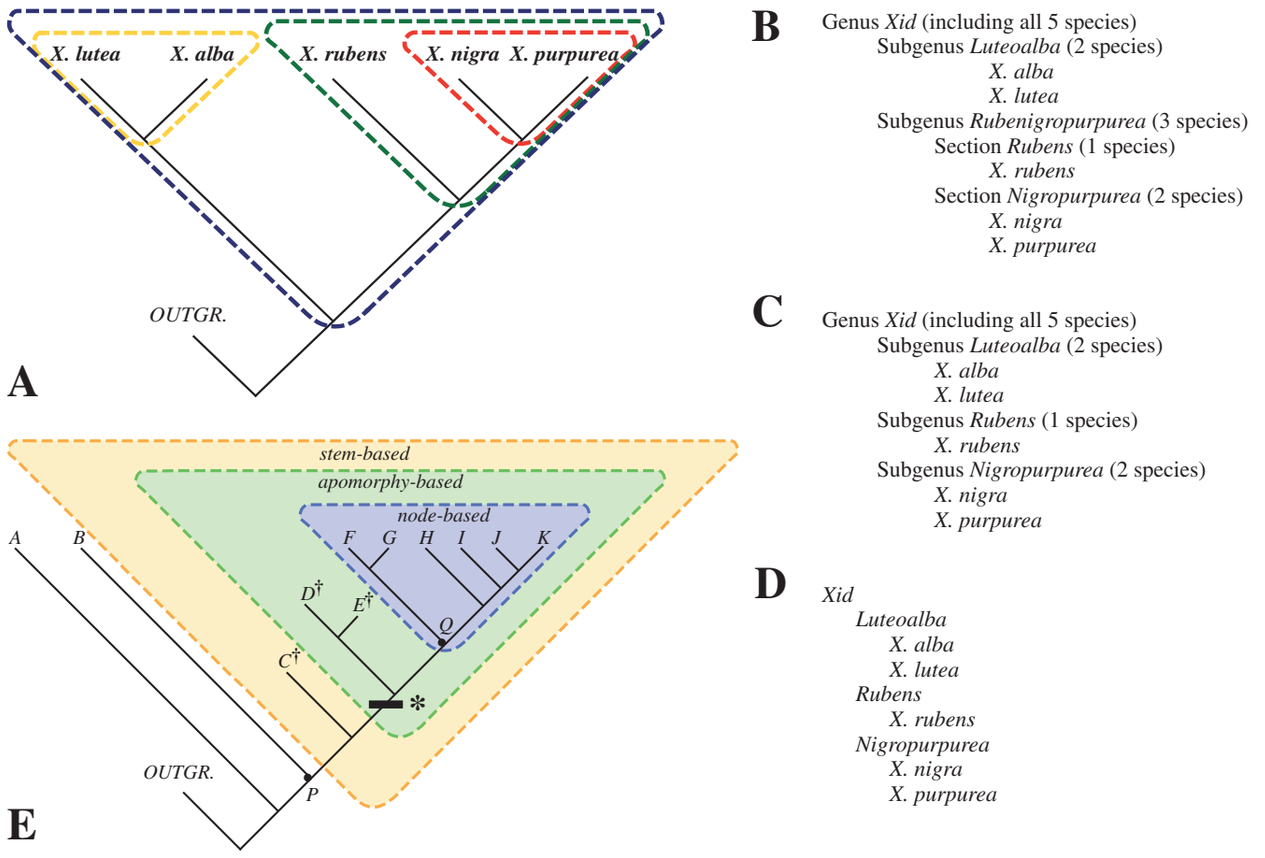


FIGURE 2.19 **A.** Cladogram from Fig. 2.6E. **B–D.** Classification schemes based on cladogram A. **B.** Indented classification. **C.** Annotated classification. **D.** Annotated, but rankless, classification. **E.** Cladogram illustrating node-based, apomorphy-based, and stem-based classification. †Extinct taxon. *Major evolutionary change, used as the basis for an apomorphy-based group.

intergradation, homology, or polarity. Thus, cladogram construction should be viewed not only as an end in itself, but as a means of pointing out those areas where additional research is needed to resolve satisfactorily the phylogeny of a group of organisms.

Cladograms represent an estimate of the pattern of evolutionary descent, both in terms of recency of common ancestry and in the distribution of derived (apomorphic) character states, which represent unique evolutionary events. Once a robust cladogram is derived, the pattern of relationships and evolutionary change may be used for a variety of purposes, discussed next.

PHYLOGENETIC CLASSIFICATION

One of the most important uses of cladograms is as a basis for classification. The pattern of evolutionary history portrayed in a cladogram may be used to classify taxa phylogenetically. A phylogenetic classification may be devised by naming and ordering monophyletic groups in a sequential, hierarchical classification, sometimes termed an **indented** method. The hierarchically arranged monophyletic groups may be assigned standard taxonomic ranks. For example, for the most parsimonious cladogram of Figure 2.19A, one possible classification of hypothetical genus *Xid* is seen in Figure 2.19B. Note that in this example, each named taxon corresponds to a monophyletic group (Figure 2.19A) and that these groups are sequentially nested such that the original cladogram may be directly reconstructed from this classification system. Two taxa of the same rank (e.g., sections *Rubens* and *Nigropurpurea*) are automatically sister groups. Each higher taxon above (e.g., subgenus *Luteoalba*) would also include automatically created lower taxa (e.g., species *Xid alba* and *Xid lutea* in this case).

An alternative, and often more practical, means of deriving a classification scheme from a cladogram is by annotation. **Annotation** is the sequential listing of derivative lineages

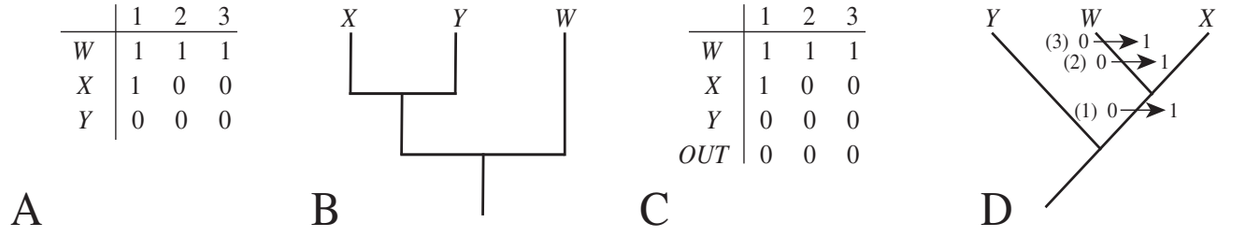


FIGURE 2.20 **A.** Character × taxon matrix for taxa *W*–*Y*. **B.** Phenogram of taxa *W*, *X*, and *Y*. **C.** Character × taxon matrix for taxa *W*–*Y* plus *OUTGROUP*. **D.** Most parsimonious cladogram of taxa *W*, *X*, and *Y*. Note different branching pattern.

from the base to the apex of the cladogram, each derivative lineage receiving the same hierarchical rank. The sequence of listing of taxa may be used to reconstruct their evolutionary relationships. For example, an annotated classification of the taxa from Figure 2.19A is seen in Figure 2.19C. In this case all named taxa are monophyletic, but taxa at the same rank are not necessarily sister groups. This is common, for example, in the naming of angiosperm orders (see Chapters 7 and 8).

The particular rank at which any given monophyletic group is given is arbitrary and is often done to conserve a past, traditional classification. A recent trend in systematics is to eliminate ranks altogether or, alternatively, to permit unranked names between the major rank names (see Chapter 16). In either case, the taxon names, minus ranks, would still retain their hierarchical, evolutionary relationship (e.g., as in Figure 2.19D).

This most common type of phylogenetic classification is sometimes termed **node-based**, because it recognizes a node (common ancestor) of the cladogram and all descendants of that common ancestor as the basis for grouping (Figure 2.19E). A node-based classification may specify a **crown clade**, one in which both or all branches from the common ancestor contain extant members. In some cases, it may be valuable to recognize a group that is **stem-based**, i.e., one that includes the "stem" (internode) region just above a common ancestor plus all descendants of that stem (Figure 2.19E). A stem-based group may be equivalent to a **total clade**, one that includes a crown clade plus all other taxa that share a recent common ancestor with the crown clade but not with other crown clades. A stem-based classification might be useful, for example, in including both a well-defined and corroborated node-based monophyletic group (crown clade) plus one or more fossil lineages that arise along the stem, the lineage below the crown clade. The paraphyletic stem may contain some, but not all, of the apomorphies possessed by the node-based crown clade. Yet a third general type of phylogenetic classification is **apomorphy-based**, in which all members of a monophyletic group that share a given, unique evolutionary event (illustrated by an "*" in Figure 2.19E) are grouped together. (See Cantino et al. 2007 for an explanation of phylogenetic classification.)

Last, it should be mentioned that a monophyletic group can be recognized with a phylogenetic "definition." For example, in Figure 2.19A, the monophyletic *Xid* might be "defined" as the "least inclusive monophyletic group containing the common ancestor of *X. lutea* and *X. nigra*." The rationale is that this presents a more explicit and stable means of classification of taxa. However, any given phylogenetic definition is based on some cladistic analysis. If future cladistic analyses portray a somewhat different relationship of taxa, then the phylogenetically defined groups may contain taxa that were unintended, making them less useful and less stable than more standard classifications.

As mentioned in Chapter 1, a second major type of classification is **phenetic**, in which taxa are grouped by overall similarity. This phenetic grouping may be represented in the form of a branching diagram known as a **phenogram**. For example, for the data matrix of Figure 2.20A, the resultant phenogram is seen in Figure 2.20B. In this case taxa *X* and *Y* share more similar features (state 0 of characters 2 and 3) than either does with taxon *W*; thus, *X* and *Y* are more similar and are grouped together. (Note that no outgroup is included in the matrix.) Phenetic classifications will often be quite different from phylogenetic ones because in a phenetic analysis, taxa may be grouped together by shared *ancestral* features (known as **symplesiomorphies**) as well as by shared derived character states (synapomorphies). For example, the data matrix of Figure 2.20C (identical to that of 2.20A except for the addition of an outgroup) yields the most parsimonious cladogram at Figure 2.20D, which has a different branching pattern from the phenogram of Figure 2.20B. Note that in the cladogram, taxa *W* and *X* are grouped as sister taxa because they share the derived state of character 1, which is a synapomorphy for *W* and *X*. In contrast, the phenogram of Figure 2.20B groups together taxa *X* and *Y* because they are more similar, having in common state "0" of characters 2 and 3; however, these are shared ancestral states (symplesiomorphies) and cannot be used to recognize monophyletic groups. Because many past

classification systems have been based on overall phenetic similarity, great caution should be taken in evaluating "relationship." Taxa that are most similar to one another may not, in fact, be particularly close relatives in a phylogenetic sense (i.e., by recency of common ancestry).

In summary, phylogenetic classification of taxa has the tremendous advantage of being based on and of reflecting the evolutionary history of the group in question. The International Code of Botanical Nomenclature (Chapter 16) has been used very successfully to assign taxonomic names based on the criterion of monophyly (although some problems persist that it is hoped will be addressed in future versions of the Code). Phylogenetic classifications have resulted in several name changes in some groups, but these are gradually beginning to stabilize, particularly with additional, robust molecular studies. In practice, assigning a name to every monophyletic group, whether ranked or not, is unwieldy, impractical, and unnecessary. Generally, only monophyletic groups that are well-supported (and ideally that have a well-recognized apomorphy) should be formally named, and every effort should be made to retain (or modify) former classification systems, where possible.

CHARACTER EVOLUTION

Cladograms can be used as an analytical device to evaluate the ancestral conditions at the cladogram nodes and the evolutionary change (apomorphies) occurring from one node to another. This may be done using the character × taxon matrix and a preexisting tree, one inferred, e.g., by parsimony, maximum likelihood, or Bayesian methods. The character(s) evaluated may or may not have been included in the original tree reconstruction.

A standard way to evaluate character evolution is by **parsimony optimization**. Optimization of characters refers to their representation (or "plotting") on a cladogram in the most parsimonious way, such that the minimal number of character state changes occur between nodes. This method assigns those character states at ancestral nodes that minimize the number of state changes between nodes, i.e., that minimize the tree length. (In the optimization procedure, a given type of character coding is selected, such as ordered, unordered, step matrices, weighted/scaled, etc.) For example, Figures 2.21A,B show a cladogram in which the evolution of a character is explained in two different ways, but neither of which is the most parsimonious explanation. In Figures 2.21C,D the character is optimized, showing the fewest possible number of state changes. In these last two examples, character state evolution can be optimized in either of two equally parsimonious ways (with ancestral nodes assigned a different set of states). **Acctran** (accelerated transformation) optimization hypothesizes an earlier initial state change with a later *reversal* of the same character (Figure 2.21C). **Deltran** (delayed transformation) optimization hypothesizes two later, *convergent* state changes (Figure 2.21D). Note that when alternative character optimization exists, there are nodes in the cladogram that are *equivocal*, i.e., for which the character state cannot be definitively determined. Optimization is automatically performed by computer algorithms that trace characters and character states. (See **Cladistic Computer Programs**, page 52.)

Another way of assessing character evolution is using maximum likelihood in **ancestral state reconstruction**, so called because it emphasizes determining the character condition at each ancestral node rather than changes between nodes. For a given tree and character distribution of the terminal taxa, this method calculates the maximum probability of a state at each node, using a selected model of evolution (generally the one used to construct the tree). An example of maximum likelihood ancestral state reconstruction is seen in Figure 2.21E. Note that ancestral nodes often do not have discrete states, but a probability of a given state, between 0 and 100%. As with maximum likelihood tree construction, branch lengths influence the probability of state changes in ancestral state reconstruction. A relatively long branch may introduce a higher or lower probability of a state, which would not be evident in parsimony optimization.

Assessment of character evolution often yields insight into the possible adaptive significance of a feature. For example, Figure 2.21F shows parsimony optimization of chromosome number for a given tree and state distribution. Studies of character evolution may allow detection of the correlation of character shifts, indicative of, say, a genetic or adaptive linkage. It may also give insight into past classification, e.g., as to whether a particular taxon was historically grouped by an apomorphy or plesiomorphy.

BIOGEOGRAPHY AND ECOLOGY

A phylogenetic analysis can be used to evaluate past changes in biogeographic distribution and ecological habitat. Both distribution and habitat data are considered to be "extrinsic" in nature, i.e., not determined by the genetic makeup (genome) of a taxon, and, therefore, not subject to biological evolution. Thus, data on distribution and habitat cannot be included in the data matrix of a cladistic analysis. (Note that ecological data in the simple sense of the habitat a taxon occupies, such as "desert" or "salt marsh," is extrinsic. However, the propensity or capability to survive in a particular habitat, e.g., physiological or morphological adaptations that allow survival in the desert, are intrinsic and may be used directly as characters in an analysis.) A historical analysis of extrinsic data may be accomplished by superposing the data onto an existing

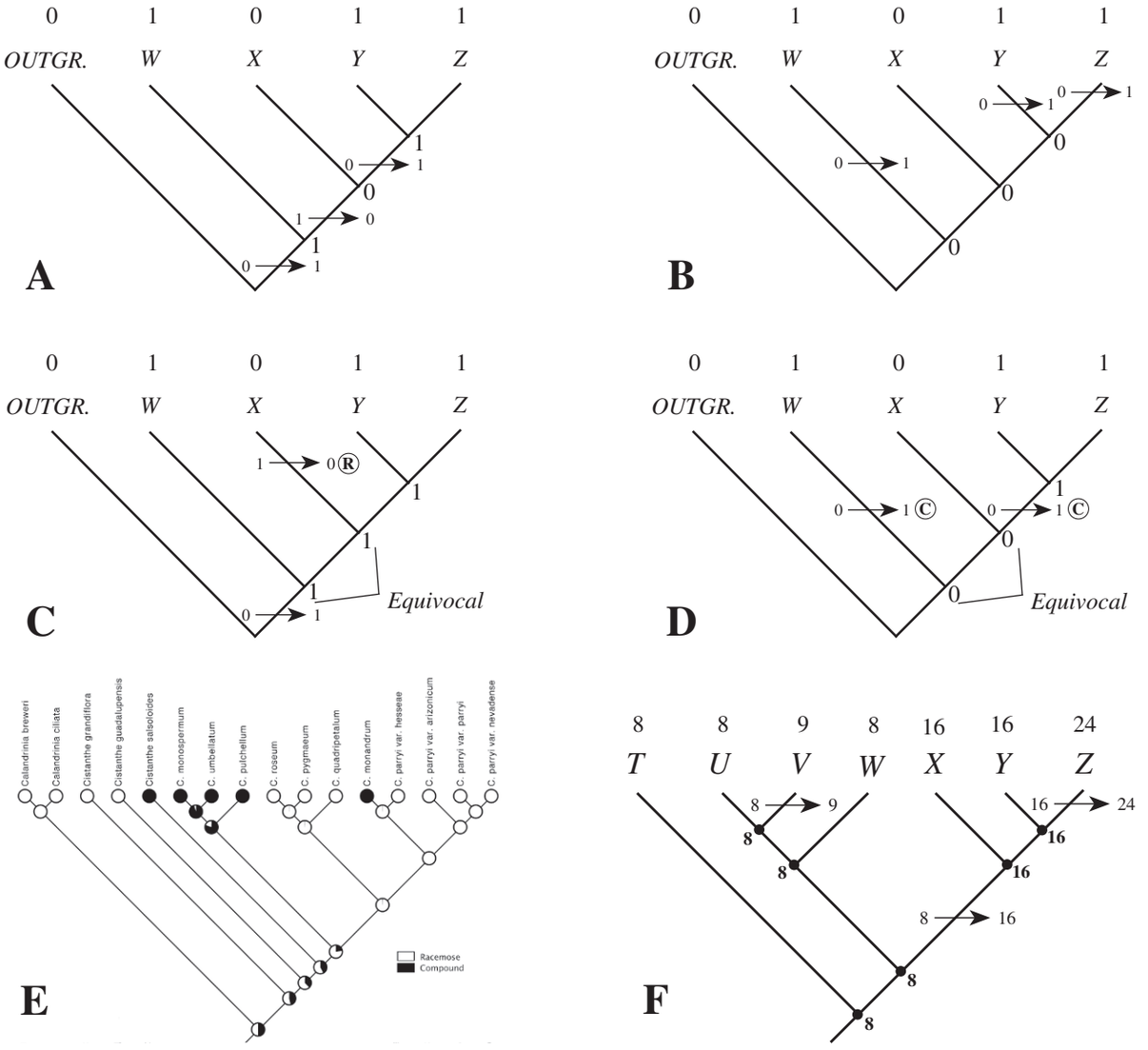


FIGURE 2.21 **A,B.** Cladograms for taxa *W–Z* and Outgroup, in which character states of a character (superposed above taxa) are accounted for by hypothesizing three state changes, not optimized. **C.** Parsimony optimization ("Acctran") of character, hypothesizing two state changes, including one reversal. **D.** Parsimony optimization ("Deltran") of character, hypothesizing two convergent state changes. **E.** Character evolution assessed by likelihood ancestral state reconstruction. Note that ancestral nodes show a probability between ca. 25% and 95% for a given state. (Example courtesy of M. Guilliams.) **F.** Character evolution using parsimony optimization, illustrated for haploid chromosome number. States at ancestral nodes in bold. Note that parsimony optimization minimizes tree length, requiring a total of three state changes.

cladogram and optimizing the changes that would be needed, e.g., using the principle of parsimony (see later discussion).

Analysis of biogeographic data can give insight into the direction of change in biogeographic distribution. A change from one distribution to another can occur by either of two means: dispersal or vicariance. **Dispersal** is the movement of an organism or propagule from one region to another, such as the transport of a seed or fruit (by wind, water, or bird) from a continent to an island (Figure 2.22A). **Vicariance**, in contrast, is the splitting of one ancestral population into two (or more) populations, e.g., by continental drift or the formation of a new waterway or mountain range, resulting in a barrier between the split populations; this barrier prevents gene flow between these populations, allowing them to diverge independently (Figure 2.22B).

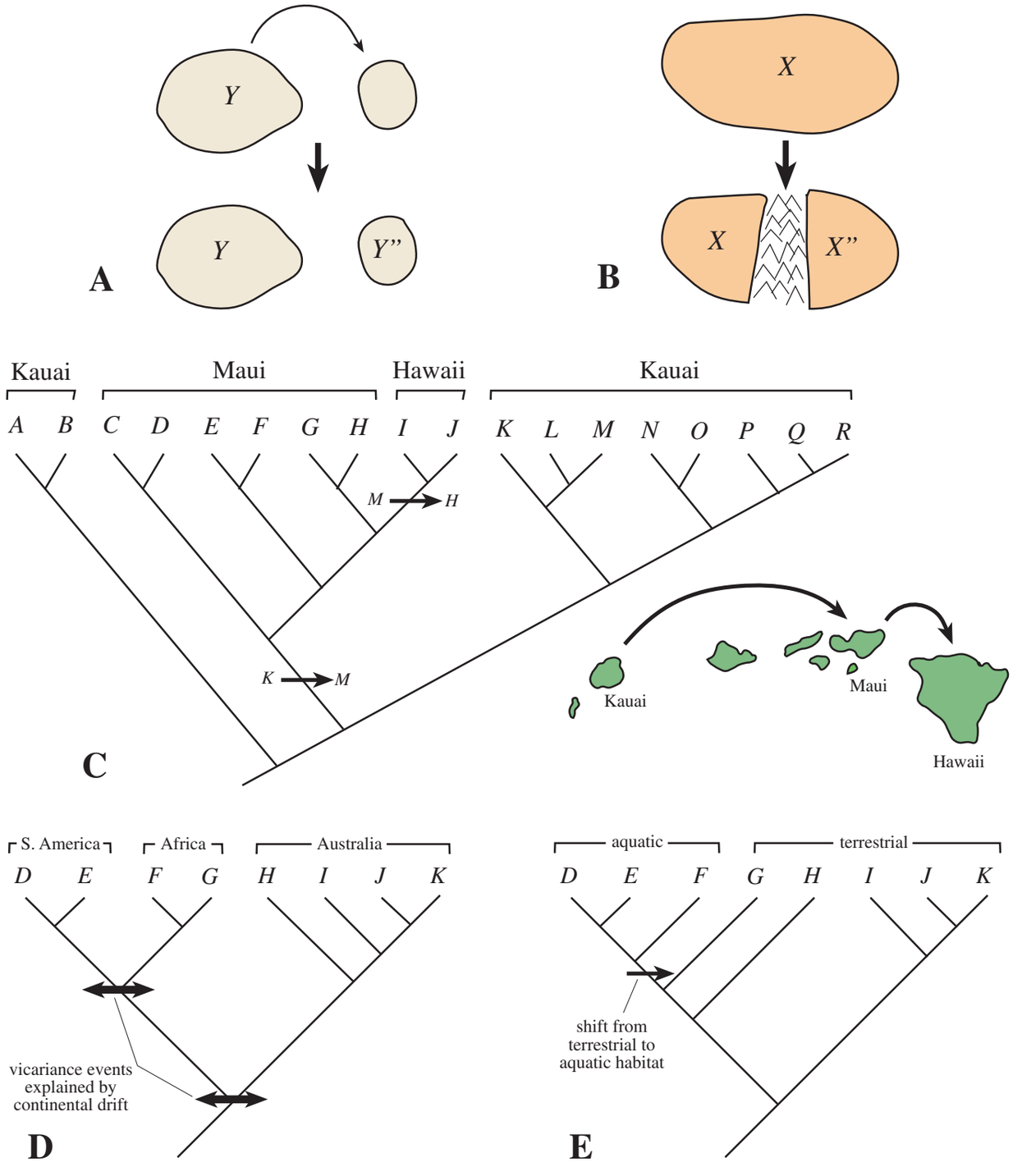


FIGURE 2.22 Cladistic analysis of biogeographic data. **A.** Hypothesis of dispersal, in which a propagule of species *Y* lands on another island. The isolated population subsequently diverges into *Y"*. **B.** Hypothesis of vicariance, in which ancestral population *X* is divided into two populations by a mountain range. The two populations, now isolated, can subsequently diverge, one becoming *X"*. **C.** Cladogram of taxa *A–R* in which geographic distributions are superposed atop lineages, illustrating dispersal in Hawaiian archipelago. Optimized explanation is dispersal of ancestral taxa from Kauai to Maui and then from Maui to Hawaii. **D.** Cladogram in which geographic distributions are superposed atop lineages, illustrating vicariance. Continental drift explains current distribution of taxa *D–K*. **E.** Superposition of ecological habitat data, illustrating use of cladogram to deduce history of ecological change.

Determining vicariance versus dispersal as an explanation for biogeographic change cannot always be made, and requires additional knowledge of geologic history. For example, Figure 2.22C illustrates a cladogram of taxa endemic to the Hawaiian archipelago, in which the ranges (by island) are superposed. A simple optimization shows the changes in geographic ranges that would be needed to explain the data. In this case, a shift from the island of Kauai to Maui and one from Maui to the island of Hawaii constitutes the simplest explanation needed to account for the current distribution of taxa. Because geologic data firmly suggests that the Hawaiian islands arose from sequential "hot-spot" volcanic activity and that the major islands were never connected, vicariance as an explanation is ruled out, leaving dispersal as the mechanism for biogeographic change. The hypothetical example of Figure 2.22D shows another cladogram in which both biogeographic distributions are superposed. A likely explanation for change in biogeographic distributions in this example is the splitting of the three continents from an ancestral Gondwana (Figure 2.22D). Although dispersal across oceans cannot be ruled out, vicariance might be more likely because the changes in distribution correspond to a hypothesis of continental drift. (Note that the continentally delimited groups need not be monophyletic.)

An example of tracing extrinsic ecological data is seen in Figure 2.22E, in which habitat types are superposed on the taxa from a cladistic analysis. Note in this example the shift from a terrestrial to an aquatic habitat. Analyses such as this may yield insight into the adaptive significance of evolutionary changes in anatomy, morphology, or physiology relative to differing habitat requirements.

ONTOGENY AND HETEROCHRONY

Phylogeny and character evolution are normally studied only with regard to the mature features of adult individuals. However, a mature structure, whether organ, tissue, or cell, is the end product of **ontogeny**, the developmental sequence under the control of a number of genes. Ontogeny may be visualized in either of two ways. First, a study of the developmental pattern may reveal a series of discrete structural stages or entities, one transforming into the next until the end point (the mature adult structure) is obtained. These discrete stages are identified and named and the transformation in **ontogenetic sequence**, from one stage to the next, is compared in different taxa (Figure 2.23A). Second, some feature of the developmental change of a structure may be measured quantitatively as a function of real time. This plot of morphology as a function of time is called an **ontogenetic trajectory** (Figure 2.23B). Ontogenetic trajectories may be compared between different taxa. Note, e.g., in Figure 2.23B that taxon *Z* and taxa *W* and *Y* have the same adult structures but differing ontogenetic trajectories.

Ontogenetic data may be used in a cladistic analysis like any other character. Thus, two or more discrete ontogenetic sequences (Figure 2.23A) or ontogenetic trajectories (Figure 2.23B) may be defined as separate character states of a developmental character. (See Appendix 4.) The polarity of ontogenetic character states may be assessed by outgroup comparison as can be done for any other character.

Evolution may often be manifested by a change in ontogeny. An evolutionary change in the rate or timing of development is known as **heterochrony**. Heterochrony has apparently been an important evolutionary mechanism in many groups, in which the relatively simple evolutionary alteration of a regulatory gene results in often profound changes in the morphology of a descendant. Heterochrony can be assessed by performing a cladistic analysis and determining from this the ancestral versus the derived condition of an ontogenetic sequence or trajectory. There are two primary categories of heterochrony, namely peramorphosis and paedomorphosis. **Peramorphosis** is a derived type of heterochrony in which ontogeny passes through *and goes beyond* the stages or trajectory of the ancestral condition. Peramorphosis can result in the addition of a new stage or an ontogenetic trajectory that continues beyond that of the ancestral trajectory. For example, in Figure 2.23C, the derived ontogenetic sequence of taxa *A* and *D* ($s^1 \Rightarrow s^2 \Rightarrow S^3$) is the result of peramorphosis via the terminal addition of stage S^3 to the ancestral sequence ($s^1 \Rightarrow S^2$). (Note that "s" represents a juvenile developmental stage; "S" is a mature, adult feature.) Thus, the adult condition (S^2) in the ancestral ontogeny is homologous with a juvenile condition (s^2) in the derived ontogeny of taxa *A* and *D*. This principle is termed *terminal addition* or *Haeckelian recapitulation* and is often summarized by the expression "ontogeny recapitulates phylogeny."

Paedomorphosis is a type of heterochrony in which the mature or adult stage of the derived ontogenetic sequence resembles a juvenile ontogenetic stage of the ancestral condition. (**Neotony** is one type of paedomorphosis that is caused by a *decrease* in the rate of development of a structure.) For example, in Figure 2.23D, the derived ontogenetic sequence of taxon *Z* ($s^1 \Rightarrow S^2$) is the result of paedomorphosis by the terminal loss of stage S^3 in the ancestral sequence ($s^1 \Rightarrow s^2 \Rightarrow S^3$). Thus, the adult condition (S^2) in the derived ontogeny of taxon *Z* is homologous with a juvenile condition (s^2) in the ancestral ontogeny. In a cladistic analysis paedomorphosis is portrayed as the *reversal* of a character state and can only be detected via the utilization of other characters in the analysis.

Evolutionary change may result in the modification of mature structures by affecting early developmental stages.

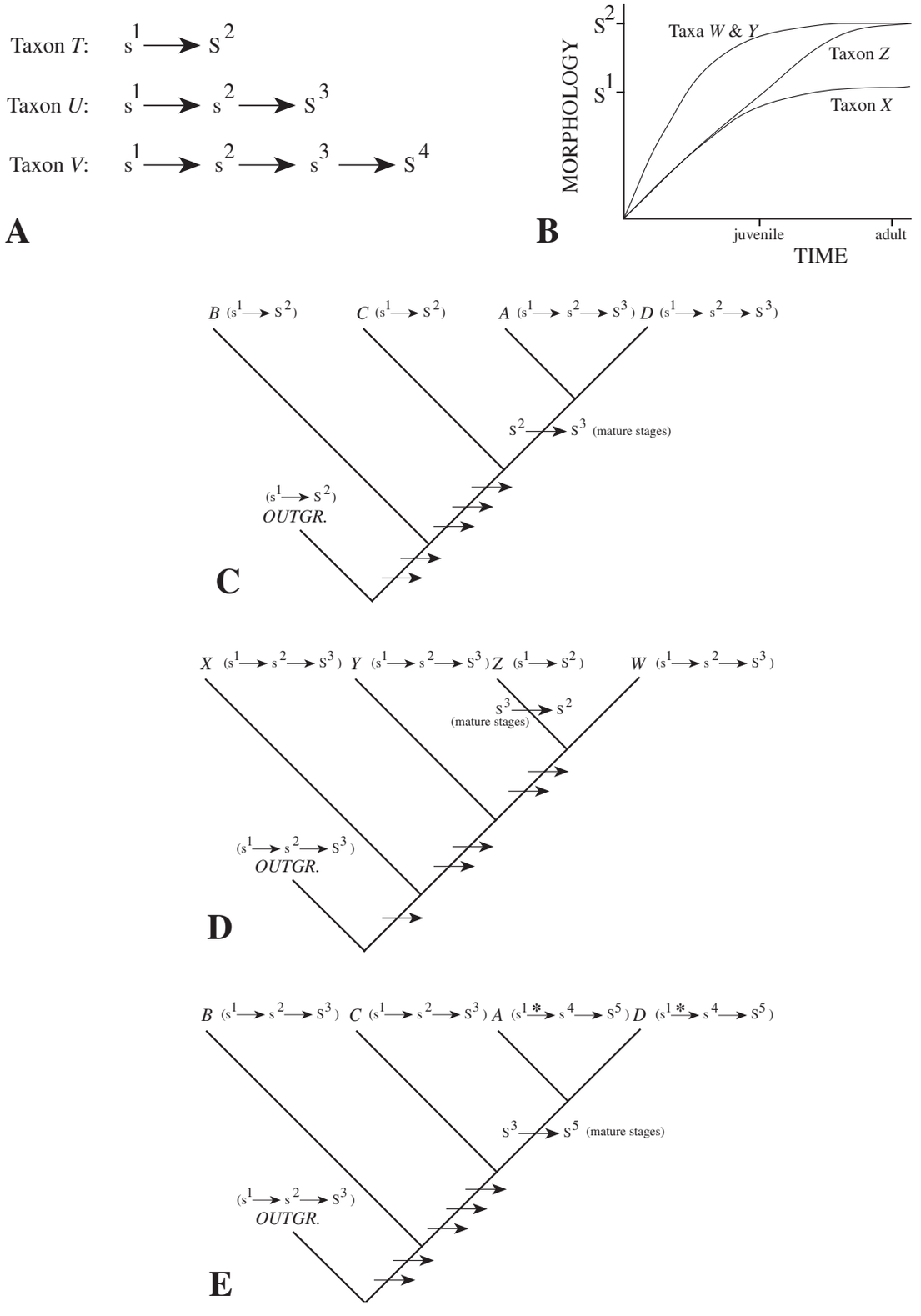


FIGURE 2.23 **A.** Representation of an ontogenetic sequence, a change from one discrete stage to another in various taxa. **B.** Ontogenetic trajectories of various taxa. Note juvenile and adult stages. **C–E.** Cladograms, with ontogenetic data (in parentheses next to taxa) and character state changes of mature structures (along lineage internodes). Note that "s" represents a juvenile developmental stage; "S" is a mature, adult feature. See text for further explanation.

For example, if the ontogeny of structure S^3 occurs in two discrete stages ($s^1 \Rightarrow s^2$) and ($s^2 \Rightarrow S^3$), then a single alteration of the regulatory pathway controlling the first developmental sequence (represented by "*" in Figure 2.23E) may cause a change in both the final structure and the intermediate stage (e.g., to $s^1 \Rightarrow s^4 \Rightarrow S^5$; Figure 2.23E). Thus, structural evolution may occur by modification at any developmental stage, and mature ancestral structures need not be preserved as extant juvenile developmental stages.

A PERSPECTIVE ON PHYLOGENETIC SYSTEMATICS

The careful researcher, in constructing cladograms or critically reading cladistic analyses in the literature, should be aware of several potential pitfalls in phylogenetic studies. Lack of consideration of any of the following renders the study questionable at best and useless at worst. Are unit taxa (OTUs) and the group as a whole monophyletic? If evidence for monophyly is not presented, the study may be faulty from the start. What are the sources of the data? The validity of a phylogenetic study is based on the comprehensiveness and accuracy of the original descriptive data. *Which* characters are selected and how are they defined? It is important to question the basis for the selection of these characters and not others. Are character states assessed for discreteness? Is homology assessed? Has an effort been made to determine whether similar characters and character states presumably have a common evolutionary origin? Or is observed similarity more one of traditional and imprecise terminology and possibly homoplasious? Have any characters been weighted? If so, what is the rationale behind it? How are polarities determined? The evidence for selection and relative placement of outgroups should be thoroughly investigated. How do cladograms compare with respect to different types of data (e.g., chloroplast versus nuclear DNA) or different types of analyses (e.g., parsimony versus Bayesian)? Finally, is the resultant cladogram analyzed in terms of monophyletic groupings, character state changes, assessment of convergences and reversals, testing of homology, and possible reevaluation of characters and character states? The thorough phylogenetic study critically reviews each step of cladogram construction, considers all alternatives, and evaluates and reevaluates the significance of the phylogenetic analysis in terms of future research that might clarify our understanding of plant evolutionary relationships. Although the determination of phylogeny using the methodology of phylogenetic systematics may be problematic, it has the significant advantage of being repeatable and explicit. Each step of the analysis can be duplicated, evaluated, and critiqued in subsequent investigations.

REVIEW QUESTIONS

OVERVIEW, TAXON SELECTION, AND CHARACTER ANALYSIS

1. Define phylogeny and give the name of the branching diagram that represents phylogeny.
2. What is phylogenetic systematics and what are its goals?
3. What are the lines of a cladogram called and what do they represent?
4. What does a split, from one lineage to two, represent?
5. Name the term for both a preexisting feature and a new feature.
6. What is the difference between an autapomorphy and a synapomorphy?
7. What does topology refer to and what is its significance in displaying cladograms?
8. What is a "phylogram" and how does it differ from a typical cladogram?
9. What names are given to both the group as a whole and the individual component taxa in a cladistic analysis?
10. What precautions must be taken in taxon selection?
11. What criteria are used in the selection and definition of characters and character states?
12. Why and how are characters assessed for character state discreteness?
13. How might characters be correlated, and what should be done in a cladistic analysis if they are?
14. What is homology and how may it be assessed?
15. What is homoplasy?
16. Name and define the two types of homoplasy and give an example of each.
17. What is a transformation series or morphocline?
18. What is character state polarity and what is the most common method for establishing polarity?
19. Name, define, and discuss the rationale for the two basic types of transformation series.

20. What is character weighting? Scaling? Why is either done?
21. What is a character step matrix? A character × taxon matrix?

CLADOGRAM CONSTRUCTION

22. What is a primary tenet of phylogenetic systematics with respect to apomorphies?
23. What is meant by recency of common ancestry?
24. What is a monophyletic group? What is the rationale for their recognition?
25. What are sister groups?
26. What is a paraphyletic group? A polyphyletic group?
27. Name a traditionally named taxonomic plant group that is not monophyletic. (Refer to Chapters 3–8.)
28. What is the principle of parsimony and what is the rationale of this principle?
29. From the data set of Figure 2.6, construct five trees that are different from the one in Figure 2.6E, draw in all character state changes, and calculate the total length of these trees. Are these trees of a different length than that of Figure 2.6E?
30. What is an unrooted tree and what can it not represent?
31. What is a polytomy and how may polytomies arise in cladistic analyses?
32. What is reticulation? How might it be detected?
33. Why do the OTUs of a study need to be verified for monophyly?
34. Why does the whole study group (ingroup) need to be verified for monophyly?
35. What is outgroup comparison and what is the rationale for using it to determine character state polarity?
36. Why should the terms ancestral/plesiomorphic and derived/apomorphic not be applied to taxa?
37. What is a consensus tree?
38. What is long branch attraction and why is it a problem in phylogenetic analysis?
39. Briefly describe the rationale and methodology of maximum likelihood. How are likelihood values calculated?
40. What are the advantages of maximum likelihood over parsimony?
41. Briefly describe the methodology of Bayesian analysis.
42. What is a consistency index and what does it measure?
43. What is a bootstrap, jackknife, decay index, and posterior probability? What do these assess?

CLADOGRAM ANALYSIS

44. Describe ways in which a classification system may be derived from a cladistic analysis.
45. What are the differences between a node-based, apomorphy-based, and stem-based classification system?
46. What is parsimony optimization and how is it used to assess character evolution?
47. How does maximum likelihood ancestral state reconstruction differ in assessing character evolution?
48. Give an example as to how a cladistic analysis can be used to assess (a) change in habitat; (b) biogeographic history.
49. Name the two major explanations for changes in distribution and indicate how they differ.
50. What is ontogeny and how may ontogeny be measured?
51. Define heterochrony, peramorphosis, paedomorphosis, and neotony.
52. Review the precautions to be taken in a cladistic analysis.
53. For the following data sets: (a) draw the three possible (dichotomously branching) cladograms; (b) for *each* of the three cladograms indicate (with arrows and corresponding characters and states) the minimum character state changes that are needed to explain the data; (c) indicate which of the three trees would be accepted by a phylogenetic systematist as the best estimate of phylogeny and why.

1

	1	2	3	4	5
A	1	1	1	1	1
B	1	0	0	0	0
C	0	0	1	1	1
OUTGROUP	0	0	0	0	0

2

	1	2	3	4	5
A	0	1	1	1	0
B	0	0	1	0	0
C	1	0	1	1	1
OUTGROUP	0	0	0	0	0

54. For each of the following data sets: (a) draw the most parsimonious cladogram; (b) indicate all character state changes; (c) circle all monophyletic groups; (d) derive a hypothetical classification scheme. Assume an ordered transformation series where more than two character states per character occur.

GENERA:	1 Flower symmetry	2 Perianth tube	3 Perianth aestivation	4 Stamen number	5 Anther shape	6 Pollen exine
Aahh	bilateral	present	valvate	6	oblong	homogeneous
Batahr	bilateral	present	valvate	6	oblong	homogeneous
Conarus	radial	present	valvate	6	oblong	homogeneous
Phlebus	radial	absent	imbricate	6	oblong	tectate
Tribus	radial	present	imbricate	6	fringed	homogeneous
OUTGROUP	radial	absent	imbricate	3	oblong	tectate

1

SPECIES:	1 Stem type	2 Carpel number	3 Pollen sculpturing	4 Perianth type	5 Staminode +/–	6 Leaf vestiture
C. cordatus	rhizome	2	psilate	rotate	–	tomentose
C. ellipticus	corm	2	psilate	rotate	–	glabrous
C. lanceolatus	rhizome	5	spinulose	rotate	–	glabrous
C. ovatus	rhizome	5	psilate	salverform	+	glabrous
C. rhomboideus	rhizome	5	psilate	salverform	+	glabrous
OUTGROUP	rhizome	5	spinulose	rotate	–	glabrous

2

GENERA:	1 Glu-Ph. allozyme	2 Pollen aperture no.	3 Anther dehiscence	4 Ovary position	5 Chromosome number	6 Leaf shape	7 Calyx merosity
Queesus	B + C	7–8	latrorse	inferior	28	ovate	5
Racamupa	B + C	7–8	latrorse	inferior	14	ovate	4
Shoota	B + C	3	poricidal	inferior	14	lanceolate	5
Tumblus	B + C	3	poricidal	inferior	14	lanceolate	5
Uvulus	A + B	3	latrorse	inferior	7	lanceolate	5
Vertex	A + B	1	latrorse	superior	7	lanceolate	4
OUTGROUP	A	1	latrorse	superior	7	lanceolate	5

3

55. Given the following data matrix and model of evolution, calculate the maximum likelihood values for at least one of the three possible unrooted trees.

Taxon	Gene Site
W	A
X	C
Y	A
Z	G

	A	C	G	T
A	0.9	0.1	0.4	0.1
C	0.1	0.9	0.1	0.4
G	0.4	0.1	0.9	0.1
T	0.1	0.4	0.1	0.9

EXERCISES

1. **Computer phylogeny applications.**
If computers are available, you may wish to explore one of the commonly used phylogeny software applications, such as MacClade (Maddison and Maddison, 2000; see others cited hereafter). These programs allow the user to input data, including taxa names and their characters and character states, and enable both the phylogenetic relationships of taxa and specific character state changes to be visualized.

With the help of your instructor, enter a data file using MacClade or some other phylogeny application for a given taxonomic group. You may use the data matrix below for the families of the Zingiberales. (Note: Root the tree at **Musaceae**.)

Examine the optimal (most parsimonious) tree. Engage the function that displays characters and visualize several, noting the distribution of their states. You may also "swap branches" on the cladogram, exploring alternative evolutionary hypotheses and noting the change in tree length.

If time allows, choose a volunteer to redraw the cladogram from MacClade onto the chalkboard. List each apomorphy illustrated on MacClade by placing the derived character state (apomorphy) beside a hatch-mark on the cladogram. Circle and tentatively name all monophyletic groups.

Review as a class the following terms: cladogram, lineage, clade, common ancestor, lineage divergence/diversification, apomorphy, synapomorphy, autapomorphy, monophyletic, paraphyletic.

Example data set of the families of the Zingiberales.

	LEAF ARRANGEMENT	SEED ARIL	POLYARC ROOT	INNER MED. STAMEN	RAPHIDES	SILICA CRYSTALS
Cannaceae	distichous	present	present	present	absent	present
Costaceae	monistichous	present	present	present	absent	present
Heliconiaceae	distichous	present	present	present	present	absent
Lowiaceae	distichous	present	absent	absent	present	absent
Marantaceae	distichous	present	present	present	absent	present
Musaceae	spiral	absent	absent	absent	present	absent
Strelitziaceae	distichous	present	present	absent	present	absent
Zingiberaceae	distichous	present	present	present	absent	present

	STAMEN NUMBER	STAMINODE PETALOID	PERISPERM	OUT. TEPALS FUSED	ANTHER TYPE
Cannaceae	1	present	present	absent	monothecal
Costaceae	1	present	present	present	bithecal
Heliconiaceae	5	absent	absent	absent	bithecal
Lowiaceae	5	absent	absent	absent	bithecal
Marantaceae	1	present	present	absent	monothecal
Musaceae	5	absent	absent	absent	bithecal
Strelitziaceae	5	absent	absent	absent	bithecal
Zingiberaceae	1	present	present	present	bithecal

2. **Web trees.**
Log onto The Tree of Life (*http://tolweb.org*), TreeBASE (*http://www.treebase.org*), Angiosperm Phylogeny website (*http://www.mobot.org/MOBOT/research/Apweb*), or a similar web page. These web pages contain up-to-date information on the relationships of organismal groups and plants, respectively. Browse through the trees illustrated on the sites and note the source of the data. Examine the apomorphies denoted at the nodes for these trees.

REFERENCES FOR FURTHER STUDY

Brooks, D. R., and D. A. McLennan. 1991. Phylogeny, Ecology, and Behavior: A Research Program in Comparative Biology. Univ. Chicago Press, Chicago.

Cantino, P. D., J. A. Doyle, S. W. Graham, W. S. Judd, R. G. Olmstead, D. E. Soltis, P. S. Soltis, and M. J. Donoghue. 2007. Towards a phylogenetic nomenclature of Tracheophyta. Taxon 56: 822–846.

Felsenstein, J. 2003. Inferring Phylogenies. Sinauer Associates. Sunderland, Massachusetts.

Gould, S. J. 1977. Ontogeny and Phylogeny. Belknap Press of Harvard University, Cambridge, Massachusetts.
Hennig, W. 1966. Phylogenetic Systematics. University of Illinois Press, Urbana.
Hillis, D. M., C. Moritz, and B. Mable (eds.). 1996. Molecular Systematics, 2nd. ed. Sinauer, Sunderland, Massachusetts.
Huelsenbeck, J. P., and J. P. Bollback. 2001. Empirical and hierarchical Bayesian estimation of ancestral states. Syst Biol 50: 351–366.
Kitching, I. J. 1998. Cladistics: The Theory and Practice of Parsimony Analysis, 2nd ed. Oxford University Press, Oxford.
Li, W. 1997. Molecular Evolution. Sinauer Associates. Sunderland, Massachusetts.
Maddison, W. P., and D. R. Maddison. 2008. MacClade 4: Analysis of Phylogeny and Character Evolution. Sinauer Associates, Sunderland, Massachusetts.
Nei, M. and S. Kumar. 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York.
Page, R. D., and E. C. Holmes. 1998. Molecular Evolution: A Phylogenetic Approach. Blackwell Science, Oxford.
Semple, C., and M. A. Steel. 2003. Phylogenetics. Oxford University Press, Oxford.
Wiley, E. O., D. Siegel-Causey, D. R. Brooks, and V. A. Funk. 1991. The Compleat Cladist: A Primer of Phylogenetic Procedures. Univ. Kansas Museum Nat. History Sp. Publ. no. 19.

CLADISTIC COMPUTER PROGRAMS

Felsenstein, J. 2009. PHYLIP (Phylogeny Inference Package).
http://evolution.gs.washington.edu/phylip.html
Goloboff, P. 1993. Nona. (NoName)
http://www.cladistics.com/aboutNona.htm
Huelsenbeck, J. P., and F. Ronquist. 2001+. MrBayes: Bayesian inference of phylogeny.
www.mrbayes.net
Maddison, W. P., and D. R. Maddison. 2008. MacClade 4: Analysis of phylogeny and character evolution. Sinauer Associates, Sunderland, Massachusetts. *http://macclade.org*
Maddison, W. P., and D. R. Maddison. 2009. Mesquite. A program for examining, among other things, ancestral state reconstruction. *http://mesquiteproject.org*
Nixon, K. C. 2000. WinClada. Software and documentation by the author. Cornell University, Ithaca, New York.
http://www.cladistics.com/aboutWinc.htm
Swofford, D. L. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, Massachusetts. *http://paup.csit.fsu.edu*

WEB SITE

Phylogeny Programs (J. Felsenstein). *http://evolution.genetics.washington.edu/phylip/software.html*
An exhaustive list and links to just about every phylogeny computer program available, a very few of which are listed above.